GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2004, 17:34:08 ; Search time 41 Seconds
(without alignments)
4701.593 Million cell updates/sec

Title: US-10-066-269-58
Perfect score: 4029
Sequence: 1 MGRGPWDAGPSRRLLPLLLL.....VGLGSLAVVIILTFLSFSAV 747

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3300	81.9	620	4	Q8WX60	Q8wx60 homo sapien
2	3002	74.5	570	4	Q9NTA7	Q9nta7 homo sapien
3	2957	73.4	570	4	Q96DM9	Q96dm9 homo sapien
4	2121	52.6	409	4	Q9H6V2	Q9h6v2 homo sapien
5	1792	44.5	345	4	Q8WU83	Q8wu83 homo sapien
6	335.5	8.3	662	11	Q9ESC0	Q9esc0 rattus norv
7	335.5	8.3	738	11	Q9ESC1	Q9esc1 rattus norv
8	331.5	8.2	738	11	Q91W44	Q91w44 mus musculu
9	331.5	8.2	739	11	Q923X1	Q923x1 mus musculu
10	306.5	7.6	3567	11	Q9ES77	Q9es77 mus musculu
11	299	7.4	937	5	Q9BLJ1	Q9blj1 ciona intes
12	294.5	7.3	3857	11	O88840	O88840 mus musculu
13	291	7.2	2189	5	Q9BI05	Q9bi05 eimeria ten
14	290.5	7.2	2872	11	Q9WUH8	Q9wuh8 rattus norv
15	285.5	7.1	1587	4	O00508	O00508 homo sapien
16	284.5	7.1	1511	4	O75412	O75412 homo sapien
17	284	7.0	835	4	O00718	O00718 homo sapien
18	283	7.0	611	4	Q8IYG0	Q8iyg0 homo sapien
19	283	7.0	1062	11	Q60789	Q60789 mus musculu
20	282	7.0	837	4	Q8NG72	Q8ng72 homo sapien
21	282	7.0	999	4	Q9NQ36	Q9nq36 homo sapien
22	281.5	7.0	1398	13	Q8AXM6	Q8axm6 xenopus lae
23	281.5	7.0	1399	13	Q8JFZ4	Q8jfz4 xenopus lae
24	281	7.0	1600	11	Q8K4G0	Q8k4g0 mus musculu
25	281	7.0	1666	11	Q8K4G1	Q8k4g1 mus musculu
26	279	6.9	2906	11	Q9WUH9	Q9wuh9 rattus norv
27	277.5	6.9	929	4	Q8NGA7	Q8nga7 homo sapien
28	274.5	6.8	1833	11	O08999	O08999 mus musculu
29	274	6.8	1106	5	Q17494	Q17494 caenorhabdi
30	274	6.8	1282	4	Q8TER0	Q8ter0 homo sapien
31	274	6.8	1972	5	Q8WPL1	Q8wpl1 oikopleura
32	272.5	6.8	1821	4	Q14767	Q14767 homo sapien
33	271.5	6.7	997	11	Q9JJS0	Q9jjs0 mus musculu
34	269.5	6.7	2809	4	Q96JP8	Q96jp8 homo sapien
35	268.5	6.7	708	13	P87363	P87363 gallus gall
36	267	6.6	642	11	Q62285	Q62285 mus musculu
37	267	6.6	642	11	Q91X17	Q91x17 mus musculu
38	266	6.6	823	4	Q9UHX3	Q9uhx3 homo sapien
39	266	6.6	837	4	Q8NG96	Q8ng96 homo sapien
40	266	6.6	1764	11	O35806	O35806 rattus norv
41	265.5	6.6	1963	6	Q28019	Q28019 bos taurus
42	263	6.5	1289	5	Q8SSS3	Q8sss3 dictyosteli
43	262.5	6.5	576	4	Q9Y3V7	Q9y3v7 homo sapien
44	262.5	6.5	1231	4	Q8IUI1	Q8iui1 homo sapien
45	256	6.4	1389	11	Q8CG18	Q8cg18 mus musculu

ALIGNMENTS

RESULT 1
Q8WX60
ID Q8WX60 PRELIMINARY; PRT; 620 AA.
AC Q8WX60;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE BA4O1.1 (Novel protein) (Fragment).
GN BA4O1.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Sehra H.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL138756; CAD13445.1; -.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 2.
DR SMART; SM00032; CCP; 2.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
KW EGF-like domain.
FT NON_TER 1 1
SQ SEQUENCE 620 AA; 69174 MW; 9BB5C9817E842CFF CRC64;

Query Match 81.9%; Score 3300; DB 4; Length 620;
Best Local Similarity 100.0%; Pred. No. 1.5e-276;
Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 128 ECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPD 187
Db 1 ECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPD 60

Qy 188 GYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRH 247

```
Db        61 GYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRH 120

Qy       248 AILVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLF 307
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 AILVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLF 180

Qy       308 NDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALYPGTN 367
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 NDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALYPGTN 240

Qy       368 YTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRSRKVGSEHMYQ 427
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 YTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRSRKVGSEHMYQ 300

Qy       428 FTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTVNVTLLRSPKRHSVQITIA 487
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 FTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTVNVTLLRSPKRHSVQITIA 360

Qy       488 TPPAVKQTISNISGFNETCLRWRSIKTADMEEMYLFHIWGQRWYQKEFAQEMTFNISSSS 547
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TPPAVKQTISNISGFNETCLRWRSIKTADMEEMYLFHIWGQRWYQKEFAQEMTFNISSSS 420

Qy       548 RDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQITEPPLPEVEFFTVHRGPLPRLRLR 607
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 RDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQITEPPLPEVEFFTVHRGPLPRLRLR 480

Qy       608 KAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNASDADGYVAAELLAKDVPDDAM 667
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 KAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNASDADGYVAAELLAKDVPDDAM 540

Qy       668 EIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWNKVRRHSCAVWAQVKDSSLMLLQMAG 727
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 EIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWNKVRRHSCAVWAQVKDSSLMLLQMAG 600

Qy       728 VGLGSLAVVIILTFLSFSAV 747
            ||||||||||||||||||||
Db       601 VGLGSLAVVIILTFLSFSAV 620
```

RESULT 2
Q9NTA7
ID Q9NTA7 PRELIMINARY; PRT; 570 AA.
AC Q9NTA7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFZP761E1824.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Amygdala;
RA Bloecker H., Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,
RA Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL137432; CAB70735.1; -.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 2.
DR SMART; SM00032; CCP; 2.
KW Hypothetical protein.
FT NON_TER 1 1
SQ SEQUENCE 570 AA; 63713 MW; 66B1A4B4096C3A1B CRC64;

Query Match 74.5%; Score 3002; DB 4; Length 570;
Best Local Similarity 100.0%; Pred. No. 7.8e-251;
Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       178 DCGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEI 237
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         1 DCGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEI 60

Qy       238 NCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEI 297
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 NCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEI 120

Qy       298 LTKINDVSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPE 357
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 LTKINDVSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPE 180

Qy       358 VCLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRS 417
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 VCLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRS 240

Qy       418 RKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTVNVTLLRSP 477
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 RKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTVNVTLLRSP 300

Qy       478 KRHSVQITIATPPAVKQTISNISGFNETCLRWRSIKTADMEEMYLFHIWGQRWYQKEFAQ 537
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 KRHSVQITIATPPAVKQTISNISGFNETCLRWRSIKTADMEEMYLFHIWGQRWYQKEFAQ 360

Qy       538 EMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQITEPPLPEVEFFTVH 597
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 EMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQITEPPLPEVEFFTVH 420

Qy       598 RGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNASDADGYVAAEL 657
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 RGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNASDADGYVAAEL 480

Qy       658 LAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWNKVRRHSCAVWAQVKD 717
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 LAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWNKVRRHSCAVWAQVKD 540

Qy       718 SSLMLLQMAGVGLGSLAVVIILTFLSFSAV 747
            |||||||||||||||||||||||||||||||
Db       541 SSLMLLQMAGVGLGSLAVVIILTFLSFSAV 570
```

RESULT 3
Q96DM9
ID Q96DM9 PRELIMINARY; PRT; 570 AA.
AC Q96DM9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ32142 (C4BP).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AK056704; BAB71259.1; -.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR Pfam; PF00084; sushi; 2.
DR SMART; SM00032; CCP; 2.
KW Hypothetical protein.
SQ SEQUENCE 570 AA; 64211 MW; E67489C3A38E868A CRC64;

Query Match 73.4%; Score 2957; DB 4; Length 570;
Best Local Similarity 98.9%; Pred. No. 6.1e-247;
Matches 554; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
Qy        153 MDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFS 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFS 60

Qy        213 VPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFES 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFES 120

Qy        273 PGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVRWQINSRRINPKISYVISIK 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVRWQINSRRINPKISYVISIK 180

Qy        333 GQRLDPMESVREETVNLTTDSRTPEVCLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEV 392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GQRLDPMESVREETVNLTTDSRTPEVCLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEV 240

Qy        393 DLLEDDGSFNISIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTRE 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DLLEDDGSFNISIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTRE 300

Qy        453 QVPVVCLDLYPTTDYTVNVTLLRSPKRHSVQITIATPPAVKQTISNISGFNETCLRWRSI 512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QVPVVCLDLYPTTDYTVNVTLLRSPKRHSVQITIATPPAVKQTISNISGFNETCLRWRSI 360

Qy        513 KTADMEEMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSE 572
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KTADMEEMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSE 420

Qy        573 LPVVISLTTQITEPPLPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTF 632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LPVVISLTTQITEPPLPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTF 480

Qy        633 SCDSEGASSFFSNASDADGYVAAELLAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYC 692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 SCDSEGASSFFSNASDADGYVAAELLAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYC 540

Qy        693 IILRITSEWNKVRRHSCAVW 712
              ||||||||||:|   |  |
Db        541 IILRITSEWNKIRHSCCCRW 560
```

```
RESULT 4
Q9H6V2
ID   Q9H6V2      PRELIMINARY;      PRT;   409 AA.
AC   Q9H6V2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein FLJ21833.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA   Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA   Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK025486; BAB15149.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   409 AA;  46012 MW;  B59E76A0DA1A10DA CRC64;

  Query Match             52.6%;  Score 2121;  DB 4;  Length 409;
  Best Local Similarity   100.0%;  Pred. No. 7.8e-175;
  Matches  409;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        339 MESVREETVNLTTDSRTPEVCLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDD 398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MESVREETVNLTTDSRTPEVCLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDD 60
```

```
Qy        399 GSFNISIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVC 458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GSFNISIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVC 120

Qy        459 LDLYPTTDYTVNVTLLRSPKRHSVQITIATPPAVKQTISNISGFNETCLRWRSIKTADME 518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 LDLYPTTDYTVNVTLLRSPKRHSVQITIATPPAVKQTISNISGFNETCLRWRSIKTADME 180

Qy        519 EMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVIS 578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVIS 240

Qy        579 LTTQITEPPLPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEG 638
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LTTQITEPPLPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEG 300

Qy        639 ASSFFSNASDADGYVAAELLAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYCIILRIT 698
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ASSFFSNASDADGYVAAELLAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYCIILRIT 360

Qy        699 SEWNKVRRHSCAVWAQVKDSSLMLLQMAGVGLGSLAVVIILTFLSFSAV 747
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 SEWNKVRRHSCAVWAQVKDSSLMLLQMAGVGLGSLAVVIILTFLSFSAV 409
```

```
RESULT 5
Q8WU83
ID   Q8WU83      PRELIMINARY;      PRT;   345 AA.
AC   Q8WU83;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC021125; AAH21125.1; -.
KW   Hypothetical protein.
FT   NON_TER       1       1
SQ   SEQUENCE   345 AA;  39009 MW;  EBA1E3711BEB95B9 CRC64;

  Query Match             44.5%;  Score 1792;  DB 4;  Length 345;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-146;
  Matches  345;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        403 ISIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLY 462
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ISIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLY 60

Qy        463 PTTDYTVNVTLLRSPKRHSVQITIATPPAVKQTISNISGFNETCLRWRSIKTADMEEMYL 522
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PTTDYTVNVTLLRSPKRHSVQITIATPPAVKQTISNISGFNETCLRWRSIKTADMEEMYL 120

Qy        523 FHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQ 582
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 FHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQ 180

Qy        583 ITEPPLPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSF 642
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ITEPPLPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSF 240

Qy        643 FSNASDADGYVAAELLAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWN 702
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 FSNASDADGYVAAELLAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWN 300

Qy        703 KVRRHSCAVWAQVKDSSLMLLQMAGVGLGSLAVVIILTFLSFSAV 747
              |||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 KVRRHSCAVWAQVKDSSLMLLQMAGVGLGSLAVVIILTFLSFSAV 345


RESULT 6
Q9ESC0
ID   Q9ESC0      PRELIMINARY;      PRT;   662 AA.
AC   Q9ESC0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ETL protein.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21264911; PubMed=11050079;
RA   Nechiporuk T., Urness L.D., Keating M.T.;
RT   "ETL, a novel seven-transmembrane receptor that is developmentally
RT   regulated in the heart. ETL is a member of the secretin family and
RT   belongs to the epidermal growth factor-seven-transmembrane
RT   subfamily.";
RL   J. Biol. Chem. 276:4150-4157(2001).
DR   EMBL; AF192402; AAG33020.1; -.
DR   HSSP; P35555; 1EMN.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   EGF-like domain.
SQ   SEQUENCE   662 AA;  73803 MW;  0FED646D020F6F30 CRC64;

  Query Match             8.3%;  Score 335.5;  DB 11;  Length 662;
  Best Local Similarity   38.7%;  Pred. No. 4.7e-20;
  Matches   67;  Conservative   22;  Mismatches   67;  Indels   17;  Gaps    5;

Qy         17 LLLLLGLARGAAGAPGPDGLDVCAT-CHEHATCQQREGKKICICNYGFVGNGRTQCVDKN 75
              ||||:||:                | | |  :| |:  :    | |: |: ||| | | | :
Db          4 LLLLVGLST----LLNHSYTQNCKTPCLPNAKCEVLDEVAACFCSTGYTGNGITICEDVD 59

Qy         76 ECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLC 135
              ||    | |||:|  | || ||| | |:|||: :    | ||||::| |:|||  : :|
Db         60 EC--NETSVCGDHAVCENTNGGFSCFCVEGYQTSTGKTQFTPNDGSYCQDVDECNETSVC 117

Qy        136 RHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEI---------DC 179
                | ||:| | |:|::||    |   | |  | : | ||          ||
Db        118 GDHAVCENTNGGFSCFCVEGYQTSTGKTQFTP-NDGSYCQEIVNSNCHLEHDC 169


RESULT 7
Q9ESC1
ID   Q9ESC1      PRELIMINARY;      PRT;   738 AA.
AC   Q9ESC1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ETL protein.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21264911; PubMed=11050079;
RA   Nechiporuk T., Urness L.D., Keating M.T.;
RT   "ETL, a novel seven-transmembrane receptor that is developmentally
RT   regulated in the heart. ETL is a member of the secretin family and
RT   belongs to the epidermal growth factor-seven-transmembrane
RT   subfamily.";
RL   J. Biol. Chem. 276:4150-4157(2001).
DR   EMBL; AF192401; AAG33019.1; -.
DR   HSSP; P35555; 1EMN.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF01825; GPS; 1.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   EGF-like domain.
SQ   SEQUENCE   738 AA;  82450 MW;  680D3731FE60EF95 CRC64;

  Query Match             8.3%;  Score 335.5;  DB 11;  Length 738;
  Best Local Similarity   38.7%;  Pred. No. 5.5e-20;
  Matches   67;  Conservative   22;  Mismatches   67;  Indels   17;  Gaps    5;

Qy         17 LLLLLGLARGAAGAPGPDGLDVCAT-CHEHATCQQREGKKICICNYGFVGNGRTQCVDKN 75
              ||||:||:                | | |  :| |:  :    | |: |: ||| | | | :
Db          4 LLLLVGLST----LLNHSYTQNCKTPCLPNAKCEVLDEVAACFCSTGYTGNGITICEDVD 59

Qy         76 ECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLC 135
              ||    | |||:|  | || ||| | |:|||: :    | ||||::| |:|||  : :|
Db         60 EC--NETSVCGDHAVCENTNGGFSCFCVEGYQTSTGKTQFTPNDGSYCQDVDECNETSVC 117

Qy        136 RHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEI---------DC 179
                | ||:| | |:|::||    |   | |  | : | ||          ||
Db        118 GDHAVCENTNGGFSCFCVEGYQTSTGKTQFTP-NDGSYCQEIVNSNCHLEHDC 169


RESULT 8
Q91W44
ID   Q91W44      PRELIMINARY;      PRT;   738 AA.
AC   Q91W44;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 82.1 kDa protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye, and Retina;
RA   Strausberg R.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC017134; AAH17134.1; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
```

```
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   Hypothetical protein; EGF-like domain.
SQ   SEQUENCE   738 AA;  82108 MW;  15D2A806C239C8CB CRC64;

  Query Match             8.2%;  Score 331.5;  DB 11;  Length 738;
  Best Local Similarity   38.4%;  Pred. No. 1.2e-19;
  Matches   66;  Conservative   18;  Mismatches   65;  Indels   23;  Gaps    4;

Qy         13 RLLPLLLLLGLARGAAGAPGPDGLDVCA---------TCHEHATCQQREGKKICICNYGFV 64
              ||||||:           |   |  |:        ||  :| |:   | : | |: |:
Db          2 RLLPLLV-----------GFSTLLNCSYTQNCSKTTCLPNAKCEVHNGVEACFCSQGYS 49

Qy         65 GNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCT 124
              ||| | | | :||    : |||:|  | |  ||| | | |||:       | ||||::|
Db         50 GNGVTICEDIDEC--SESSVCGDHAVCENVNGGFSCFCREGYQTATGKSQFTPNDGSYCQ 107

Qy        125 DIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTE 176
              |||||  | :|     | | :| | |:| :||    |   | |  | : | |
Db        108 DIDECSESSVCGDHAVCENVNGGFSCFCREGYQTATGKSQFTP-NDGSYCQE 158


RESULT 9
Q923X1
ID   Q923X1      PRELIMINARY;      PRT;   739 AA.
AC   Q923X1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   ETL1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AKR/J;
RX   MEDLINE=21332331; PubMed=11438738;
RA   Terskikh A.V., Easterday M.C., Li L., Hood L., Kornblum H.I.,
RA   Geschwind D.H., Weissman I.L.;
RT   "From hematopoiesis to neuropoiesis: Evidence of overlapping genetic
RT   programs.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7934-7939(2001).
DR   EMBL; AF385682; AAK62363.1; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 2.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   EGF-like domain.
SQ   SEQUENCE   739 AA;  82281 MW;  0A41AB8A524E98D7 CRC64;

  Query Match             8.2%;  Score 331.5;  DB 11;  Length 739;
  Best Local Similarity   38.4%;  Pred. No. 1.2e-19;
  Matches   66;  Conservative   18;  Mismatches   65;  Indels   23;  Gaps    4;

Qy         13 RLLPLLLLLGLARGAAGAPGPDGLDVCA---------TCHEHATCQQREGKKICICNYGFV 64
              ||||||:           |   |  |:        ||  :| |:   | : | |: |:
Db          2 RLLPLLV-----------GFSTLLNCSYTQNCSKTTCLPNAKCEVHNGVEACFCSQGYS 49

Qy         65 GNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCT 124
              ||| | | | :||    : |||:|  | |  ||| | | |||:       | ||||::|
Db         50 GNGVTICEDIDEC--SESSVCGDHAVCENVNGGFSCFCREGYQTATGKSQFTPNDGSYCQ 107

Qy        125 DIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTE 176
              |||||  | :|     | | :| | |:| :||    |   | |  | : | |
Db        108 DIDECSESSVCGDHAVCENVNGGFSCFCREGYQTATGKSQFTP-NDGSYCQE 158


RESULT 10
Q9ES77
ID   Q9ES77      PRELIMINARY;      PRT;  3567 AA.
AC   Q9ES77;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Polydom protein precursor.
GN   POLYDOM.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C3H/HeNSIc;
RX   MEDLINE=20517255; PubMed=11062057;
RA   Gilges D., Vinit M.-A., Callebaut I., Coulombel L., Cacheux V.,
RA   Romeo P.-H., Vigon I.;
RT   "Polydom : a secreted protein with pentraxin, complement control
RT   protein, epidermal growth factor and von willebrand factor A
RT   domains.";
RL   Biochem. J. 352:49-59(2000).
DR   EMBL; AF206329; AAG32160.1; -.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:1928849; Polydom.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR003410; Hyalin.
DR   InterPro; IPR001759; Pentaxin.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00008; EGF; 10.
DR   Pfam; PF02494; HYR; 2.
DR   Pfam; PF00354; pentaxin; 1.
DR   Pfam; PF00084; sushi; 33.
DR   Pfam; PF00092; vwa; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR00895; PENTAXIN.
DR   ProDom; PD002153; Pentaxin; 1.
DR   SMART; SM00032; CCP; 34.
DR   SMART; SM00179; EGF_CA; 6.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 6.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 6.
DR   PROSITE; PS50234; VWFA; 1.
KW   EGF-like domain; Signal.
FT   SIGNAL       1     17       POTENTIAL.
SQ   SEQUENCE   3567 AA;  387391 MW;  8FBA8276E12293E5 CRC64;

  Query Match             7.6%;  Score 306.5;  DB 11;  Length 3567;
  Best Local Similarity   20.6%;  Pred. No. 1.8e-16;
  Matches  102;  Conservative   69;  Mismatches  152;  Indels  171;  Gaps   20;

Qy         46 ATCQQREGKKICI-CNYGF--VGNGRTQCVDKNE-------CQFGATLVCGNHTSCHNTP 95
              |:  :: | |: : |: ||  |||    |::: :       |:    : ||   :  |
Db       1643 ASGNRKPGSKVSLFCDPGFQMVGNPVQYCLNQGQWTQPLPHCE---RIRCGLPPALEN-- 1697
```

```
Qy         96 GGFYCI------------CLEGYRATNNNK-FIPNDGTF------CTDIDECEVSGLCRH 137
              |||              |  ||    :::  |  ::|::       | |:||| |   |
Db       1698 -GFYSAEDFHAGSTVTYQCTSGYYLLGDSR-FCTDNGSWNGISPSCLDVDECAVGSDCSE 1756

Qy        138 GGRCVNTHGSFECYCMDGY--------------LPRNGPEPFHPTTD------------- 170
               |:||:||: | |   |              | | ||  | : :
Db       1757 HASCLNTNGSYVCSCNPPYTGDGKNCAEPVKCKAPEN-PENGHSSGEIYTVGTAVTFSCD 1815

Qy        171 ----------------------ATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACRE 208
                                   ||  | || || ||:   :     : ||:| | | :
Db       1816 EGHELVGVSTITCLETGEWDRLRPSCEAISCGVPP-VPENGGVDGSAFTYGSKVVYRCDK 1874

Qy        209 GF--------------------------FSVPED------------------------- 216
              |:                         | |||
Db       1875 GYTLSGDEESACLASGSWSHSSPVCELVKCSQPEDINNGKYILSGLTYLSIASYSCENGY 1934

Qy        217 -----TVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFE 271
                 ::  ||  |:|:     || ::|| || :: |:: |:: :  |    | |:||:
Db       1935 SLQGPSLLECTASGSWDRAPPSCQLVSCGEPPIVKDAVITGSNFT-FGNTVAYTCKEGY- 1992

Qy        272 SPGGKITSVCTEKGTWRESTLTCTEIL----TKINDVS------LFNDTCVRWQINSRRI 321
               :  |  | ||   | |  |   |  :       ::  |      || ||   :  :   :
Db       1993 TLAGPDTIVCQANGKWNSSNHQCLAVSCDEPPNVDHASPETAHRLFGDTAFYYCADGYSL 2052

Qy        322 NPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALY----PGTNYTVNISTAPP 377
                 : : :|  : |                  |  |:| :    |  :|::  | :
Db       2053 ADNSQLICNAQGNWVPP-----------AQAVPR-CIAHFCEKPPSVSYSILESVSKA 2099

Qy        378 RRSMPAVIGFQTAE 391
               : :  :|: |:  |
Db       2100 KFAAGSVVSFKCME 2113
```

RESULT 11
Q9BLJ1
ID Q9BLJ1 PRELIMINARY; PRT; 937 AA.
AC Q9BLJ1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Ci-META1.
GN CI-META1.
OS Ciona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
OC Phlebobranchia; Cionidae; Ciona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21347414; PubMed=11455433;
RA Nakayama A., Satou Y., Satoh N.;
RT "Isolation and characterization of genes that are expressed during
RT Ciona intestinalis metamorphosis.";
RL Dev. Genes Evol. 211:184-189(2001).
DR EMBL; AB041857; BAB40596.1; -.
DR HSSP; P00742; 1HCG.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR001212; Somatomedin_B.
DR InterPro; IPR001491; Thrmbomoduln.
DR Pfam; PF00008; EGF; 16.
DR PRINTS; PR00907; THRMBOMODULN.
DR SMART; SM00179; EGF_CA; 14.
DR PROSITE; PS00010; ASX_HYDROXYL; 14.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 19.
DR PROSITE; PS01187; EGF_CA; 12.
DR PROSITE; PS00524; SOMATOMEDIN_B; 1.
KW EGF-like domain.
SQ SEQUENCE 937 AA; 101043 MW; 8C67830C8E391D07 CRC64;

Query Match 7.4%; Score 299; DB 5; Length 937;
Best Local Similarity 26.2%; Pred. No. 1.1e-16;
Matches 95; Conservative 31; Mismatches 125; Indels 112; Gaps 17;

```
Qy         26 GAAGAP--GPDGLDVC--ATCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGA 81
              |  | |  |   :| |  : |  ||||     : :| |  |: || :| | | |||
Db       190 GYTGNPQTGCTAIDPCNPSPCGAHATCSSVHYRPVCTCEAGYTGNPQTGCTDVNECL--R 247

Qy         82 TLVCG---NHTSCHNTPGGFYCICLEGYRA-----------------TNNNKTFIPNDG 120
                |||   |:  | || | | |:|  ||||                    |:||  :   |
Db       248 PNVCGPASNNKRCVNTAGSFRCVCKNGYRAQGSRCVDINECLRPNICGINSNKRCVNRPG 307

Qy        121 TF--------------CTDIDEC---EVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPE 163
               :              | ||:||    | |    |  ||||| ||| | | :||
Db       308 GYRCVCKPGYRAQGSRCVDINECLRPNVCGPASHNKRCVNTPGSFRCVCRNGY------- 360

Qy        164 PFHPTTDATSCTEI-DCGTPPEVPDGYIIGNYTSSLGSQVRYACR-----EGFFSVPEDT 217
                    : | :| :| : |         ||    : :   : ||       |   :  |
Db       361 ----RAQGSRCVDINECRSSP-------CGNNAQCINTPGSFTCRCNTGYTGNGRICRD- 408

Qy        218 VSSCTGLGTWESPK----------------------LHCQEIN-------CGNPPEMRHA 248
              |: |  :    ||                      | |::||       ||
Db       409 VNECAAIRPPCSPNADCTNTPGSFTCQCKPGYTGNGLVCRDINECLSRSACG-------- 460

Qy        249 ILVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFN 308
              :          | | ||: |: : | :    | :     | |    |      ||    |
Db       461 VNTNKRCINTPGSFRCVCRNGYRAQGSR----CVDINECRSS--PCGNNARCINTPGSFT 514

Qy        309 DTC 311
                |
Db       515 CRC 517
```

RESULT 12
O88840
ID O88840 PRELIMINARY; PRT; 3857 AA.
AC O88840;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mutant fibrillin-1.
GN FBN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.D2;
RX MEDLINE=98069008; PubMed=9405934;
RA Bona C.A., Murai C., Casares S., Kasturi K., Nishimura H., Honjo T.,
RA Matsuda F.;
RT "Structure of the mutant fibrillin-1 gene in the tight skin (TSK)
RT mouse.";
RL DNA Res. 4:267-271(1997).
DR EMBL; AF007248; AAC62317.1; -.
DR HSSP; P35555; 1APJ.
DR MGD; MGI:95489; Fbn1.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR002557; Chitin_bind_PerA.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002212; Fibril-assoc.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00008; EGF; 64.
DR Pfam; PF00683; TB; 12.
DR SMART; SM00494; ChtBD2; 2.
DR SMART; SM00179; EGF_CA; 60.
DR PROSITE; PS00010; ASX_HYDROXYL; 61.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 50.
DR PROSITE; PS01187; EGF_CA; 61.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.

```
KW   EGF-like domain.
SQ   SEQUENCE   3857 AA;  418301 MW;  5BC0618BC527E04C CRC64;

  Query Match             7.3%;  Score 294.5;  DB 11;  Length 3857;
  Best Local Similarity   26.7%;  Pred. No. 2.2e-15;
  Matches   83;  Conservative   33;  Mismatches   94;  Indels  101;  Gaps   15;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  |:| | | |||: ||   || |  | ||
Db       1291 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCGRHAVCTNTA 1345

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
              | | | |  |:           ||  |||:|||   : :|     | || ||: | | |
Db       1346 GSFKCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKD 1395

Qy        155 GYLPRNGPEPFHPTTDATSCTEID--------CGTPP--EVPDGYIIGNYTSSLGSQVRY 204
              ||           | |  :||::|        ||       | ||               |
Db       1396 GY-----------TGDGFTCTDLDECSENLNLCGNGQCLNAPGGY-------------RC 1431

Qy        205 ACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARY 264
              |  ||  ||     :|  :     | : |    | | |                 |: |
Db       1432 ECDMGF--VPSADGKACEDIDECSLPNI-CVFGTCHNLP---------------GLFRC 1472

Qy        265 VCQEGFE--SPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVR--------- 313
              |: |:|    ||                  ||::   ::  :  :  ||
Db       1473 ECEIGYELDRSGG----------------NCTDVNECLDPTTCISGNCVNTPGSYTCDC 1515

Qy        314 ---WQINSRRI 321
                 :::|  |:
Db       1516 PPDFELNPTRV 1526


RESULT 13
Q9BI05
ID   Q9BI05      PRELIMINARY;      PRT;  2189 AA.
AC   Q9BI05;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Microneme protein 4.
GN   MIC4.
OS   Eimeria tenella.
OC   Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC   Eimeria.
OX   NCBI_TaxID=5802;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Houghton;
RA   Tomley F.M., Billington K.J., Bumstead J.M., Clark J.D., Monaghan P.;
RT   "EtMIC4: A microneme protein from Eimeria tenella that contains tandem
RT   arrays of epidermal growth factor-like and thrombospondin type-I
RT   repeats.";
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ306453; CAC34726.1; -.
DR   HSSP; P35555; 1EMN.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 24.
DR   SMART; SM00179; EGF_CA; 22.
DR   PROSITE; PS00010; ASX_HYDROXYL; 22.
DR   PROSITE; PS01186; EGF_2; 18.
DR   PROSITE; PS01187; EGF_CA; 18.
KW   EGF-like domain.
SQ   SEQUENCE   2189 AA;  224388 MW;  C272A420B94FCB2D CRC64;

  Query Match             7.2%;  Score 291;  DB 5;  Length 2189;
  Best Local Similarity   30.8%;  Pred. No. 1.9e-15;
  Matches   94;  Conservative   27;  Mismatches   80;  Indels  104;  Gaps   22;

Qy         36 LDVCA----TCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSC 91
              :| ||     ||  |||   :|   | || || |||  :| | :||   |   || :| |
Db        600 IDECANGTHNCHASATCTNTQGSFECACNAGFSGNG-VECNDVDECSTDAD-DCGENTLC 657

Qy         92 HNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSG--LCRHGGRCVNTHGSFE 149
              :|| | | | |: |: |          |   | |||||  ||   |      | || |||
Db        658 NNTVGSFECTCMAGFEAA---------DAKTCKDIDEC-ASGTHTCSTHATCTNTAGSFT 707

Qy        150 CYC---MDGYLPRNGPEPFHPTTDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQVRYA 205
              | |     ||              |   | ::| ||           | :  :: ::
Db        708 CECNPSFDG-------------DGHKCEDVDFCGQ----------GLHDCNVHAE---- 739

Qy        206 CREGFFSVPEDTVSSCT-GLG-TWESPKLH----CQEIN------------CGNPPEMR 246
              | |       ::|    || |:| | |     |    ||:|:              | | |
Db        740 CSES----DDNTTFKCTCGIGYTGEG---HGENGCQDIDECAQDAICGENTVCTNTP--- 789

Qy        247 HAILVGNHSSRLGGVARYVCQEGFBSPGGKITSVCTEKGTWRESTLTCTEILTKINDVSL 306
                        |     | ||| : | |:      ||    ::||| :|  : || |
Db        790 ------------GSFECACVEGFVAVGAKL------KGA---TSLTCIDI-DECNDAS- 825

Qy        307 FNDTC 311
                :||
Db        826 -KNTC 829


RESULT 14
Q9WUH8
ID   Q9WUH8      PRELIMINARY;      PRT;  2872 AA.
AC   Q9WUH8;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Fibrillin-1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99032689; PubMed=9815129;
RA   Kanwar Y.S., Ota K., Yang Q., Kumar A., Wada J., Kashihara N.,
RA   Peterson D.R.;
RT   "Isolation of rat fibrillin-1 cDNA and its relevance in metanephric
RT   development.";
RL   Am. J. Physiol. 275:F710-F723(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kanwar Y.S.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF135059; AAD34438.1; -.
DR   HSSP; P35555; 1APJ.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00008; EGF; 45.
DR   Pfam; PF00683; TB; 9.
DR   SMART; SM00179; EGF_CA; 41.
DR   PROSITE; PS00010; ASX_HYDROXYL; 42.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 38.
DR   PROSITE; PS01187; EGF_CA; 41.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
KW   EGF-like domain.
SQ   SEQUENCE   2872 AA;  312069 MW;  0C4F7F3B87A80280 CRC64;

  Query Match             7.2%;  Score 290.5;  DB 11;  Length 2872;
  Best Local Similarity   25.9%;  Pred. No. 3.2e-15;
  Matches   78;  Conservative   36;  Mismatches  106;  Indels   81;  Gaps   12;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  |:| | | |||: ||   || |  | ||
Db       1290 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCGRHAVCTNTA 1344

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
```

```
Db       1345 GSFKCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKD 1394

Qy        155 GYLPRNGPEPFHPTTDATSCTEIDCGTPPE/PDGYIIGNYTSSLGSQVRYACREGFFSVP 214

Db       1395 GY-----------TGDGFTCTDLD-----E:SENLNLSGNGQCLNAPAGYRCECDMGFVP 1438

Qy        215 EDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFE--S 272

Db       1439 SADGKACEDINECSLPNI-CVFGTCHNLP----------------GLFRCECEIGYELDR 1481

Qy        273 PGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVR-----------WQINSRR 320

Db       1482 SGG----------------NCTDVNECLDPTTCISGNCVNTPGSYTCDCPPDFELNPTR 1524

Qy        321 I 321

Db       1525 V 1525
```

RESULT 15
O00508
ID O00508 PRELIMINARY; PRT; 1587 AA.
AC O00508;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Latent TGF-beta binding protein-4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97415399; PubMed=9271198;
RA Giltay R., Kostka G., Timpl R.;
RT "Sequence and expression of a novel member (LTBP-4) of the family of
RT latent transforming growth factor-beta binding proteins.";
RL FEBS Lett. 411:164-168(1997).
DR EMBL; Y13622; CAA73944.1; -.
DR HSSP; P35555; 1EMN.
DR Genew; HGNC:6717; LTBP4.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002212; Fibril-assoc.
DR InterPro; IPR003018; GAF.
DR Pfam; PF00008; EGF; 17.
DR Pfam; PF00683; TB; 4.
DR SMART; SM00179; EGF_CA; 15.
DR SMART; SM00065; GAF; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 14.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 12.
DR PROSITE; PS01187; EGF_CA; 17.
KW EGF-like domain.
SQ SEQUENCE 1587 AA; 169620 MW; 57A832F95FA0AE46 CRC64;

Query Match 7.1%; Score 285.5; DB 4; Length 1587;
Best Local Similarity 22.5%; Pred. No. 3.6e-15;
Matches 127; Conservative 50; Mismatches 180; Indels 207; Gaps 25;

```
Qy         22 GLARGAAGAPGPDGLDVCATCHEHAT---CQQREGKKICICNYGFVGN------------ 66

Db        703 GFRSRGPGAPCQD-VDECARSPPPCTYGRCENTEGSFQCVCPMGFQPNTAGSECEDVDEC 761

Qy         67 ------------------------------GRTQCVDKNECQFGATLVCGNHTSCHNTPGG 97

Db        762 ENHLACPGQECVNSPGSFQCRTCPSGHHLHRGRCTDVDECSSGAP-PCGPHGHCTNTEGS 820

Qy         98 FYCICLEGYRATNN------------------NKTFIPNDGTF---------------- 122

Db        821 FRCSCAPGYRAPSGRPGPCADVNECLEGDFCFPHGECLNTDGSFACTCAPGYRPGPRGAS 880
```

```
Qy        123 CTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEID---- 178

Db        881 CLDVDECSEEDLCQ-SGICTNTDGSFECICPPGH--RAGP-------DLASCLDVDECRE 930

Qy        179 -----CGTPPEVPDGYIIGNYTSSLGSQ--VRYACREGFFSVPE---DTVSSCTGLGTWE 228

Db        931 RGPALCGSQ----------RCENSPGSYRCVR-DCDPGYHAGPEGTCDDVDECQEYG--- 976

Qy        229 SPKLHCQEINCGNPP------------------------EMRHAILVGNHS--SRLGGV 261

Db        977 -PEI-CGAQRCENTPGSYRCTPACDPGYQPTPGGGCQDVDECRNRSFCGAHAVCQNLPGS 1034

Qy        262 ARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVRWQINSRRI 321

Db       1035 FQCLCDQGYE--GARDGRHCVD----------------VNECETLQGVC---------- 1065

Qy        322 NPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALYPGTNYTVNISTAPPRRSM 381

Db       1066 --GAALCENVEGSFLCVCPNSPEEFDPMTGRCVPPRTSVGMSPGSQPQAPVSPVLPARPP 1123

Qy        382 PAVIG----------------FQTAEVDLLEDDGSFNISIFNETCLKLN-------R 415

Db       1124 PPPLSRRPRKPRKGPVGSGCRECYFDTAAPDACDNILARNVT-WQECCCTVGEGWGSGCR 1182

Qy        416 RSRKVGSEHM-YQFTVLGQRWYLA 438

Db       1183 IQQCPGTETAEYQSLCPHGRGYLA 1206
```

Search completed: February 6, 2004, 17:39:04
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2004, 17:25:33 ; Search time 17 Seconds
(without alignments)
2066.409 Million cell updates/sec

Title: US-10-066-269-58
Perfect score: 4029
Sequence: 1 MGRGPWDAGPSRRLLPLLLL.....VGLGSLAVVIILTFLSFSAV 747

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	294.5	7.3	2871	1	FBN1_MOUSE	Q61554 mus musculu
2	290.5	7.2	2871	1	FBN1_HUMAN	P35555 homo sapien
3	287	7.1	835	1	CD97_HUMAN	P48960 homo sapien
4	283.5	7.0	2871	1	FBN1_BOVIN	P98133 bos taurus
5	281	7.0	2907	1	FBN2_MOUSE	Q61555 mus musculu
6	278.5	6.9	2911	1	FBN2_HUMAN	P35556 homo sapien
7	277.5	6.9	640	1	UROM_HUMAN	P07911 homo sapien
8	274.5	6.8	886	1	EMR1_HUMAN	Q14246 homo sapien
9	270	6.7	2871	1	FBN1_PIG	Q9tv36 sus scrofa
10	260.5	6.5	1394	1	LTBS_HUMAN	P22064 homo sapien
11	260.5	6.5	1595	1	LTBL_HUMAN	Q14766 homo sapien
12	260	6.5	931	1	EMR1_MOUSE	Q61549 mus musculu
13	256	6.4	1389	1	LTBS_MOUSE	Q8cg18 mus musculu
14	256	6.4	1713	1	LTBL_MOUSE	Q8cg19 mus musculu
15	251	6.2	1712	1	LTB1_RAT	Q00918 rattus norv
16	246.5	6.1	1247	1	NIDO_HUMAN	P14543 homo sapien
17	246	6.1	2319	1	NTC3_RAT	Q9r172 rattus norv
18	245.5	6.1	643	1	UROM_BOVIN	P48733 bos taurus
19	245.5	6.1	703	1	FBL1_HUMAN	P23142 homo sapien
20	242.5	6.0	816	1	NEL2_MOUSE	Q61220 mus musculu
21	241.5	6.0	705	1	FBL1_MOUSE	Q08879 mus musculu
22	241.5	6.0	2318	1	NTC3_MOUSE	Q61982 mus musculu
23	240	6.0	1221	1	FBL2_MOUSE	P37889 mus musculu
24	238.5	5.9	816	1	NEL2_RAT	Q62918 rattus norv
25	236.5	5.9	704	1	FBL1_CHICK	O73775 gallus gall
26	236	5.9	493	1	FBL3_HUMAN	Q12805 homo sapien
27	235	5.8	644	1	UROM_RAT	P27590 rattus norv
28	234	5.8	2201	1	TENA_HUMAN	P24821 homo sapien
29	233	5.8	570	1	FBP3_STRPU	P49013 strongyloce
30	233	5.8	2003	1	NTC4_HUMAN	Q99466 homo sapien
31	231.5	5.7	2321	1	NTC3_HUMAN	Q9um47 homo sapien
32	231	5.7	1245	1	NIDO_MOUSE	P10493 mus musculu
33	231	5.7	1964	1	NTC4_MOUSE	P31695 mus musculu
34	230	5.7	2524	1	NOTC_XENLA	P21783 xenopus lae
35	229	5.7	493	1	FBL3_RAT	O35568 rattus norv
36	229	5.7	1133	1	EGF_RAT	P07522 rattus norv
37	229	5.7	1403	1	NID2_MOUSE	O88322 mus musculu
38	228.5	5.7	816	1	NEL2_HUMAN	Q99435 homo sapien
39	228.5	5.7	816	1	NEL_CHICK	Q90827 gallus gall
40	225	5.6	1184	1	FBL2_HUMAN	P98095 homo sapien
41	222	5.5	2703	1	NOTC_DROME	P07207 drosophila
42	221	5.5	798	1	FBL1_CAEEL	O77469 caenorhabdi
43	220.5	5.5	1064	1	FBP1_STRPU	P10079 strongyloce
44	220.5	5.5	1213	1	JAG3_BRARE	Q90y54 brachydanio
45	220.5	5.5	1375	1	NID2_HUMAN	Q14112 homo sapien

ALIGNMENTS

RESULT 1
FBN1_MOUSE
ID FBN1_MOUSE STANDARD; PRT; 2871 AA.
AC Q61554; Q60826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130561; PubMed=7829516;
RA Yin W., Germiller J., Sanguineti C., Smiley E., Pangilinan T.,
RA Pereira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene.";
RL J. Biol. Chem. 270:1798-1806(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS (BY SIMILARITY).
CC -!- SIMILARITY: Contains 47 EGF-like domains.
CC -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; L29454; AAA56840.1; -.
DR EMBL; U22493; AAA64217.1; -.
DR PIR; A55624; A55624.
DR HSSP; P35555; 1APJ.
DR MGD; MGI:95489; Fbn1.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002212; Fibril-assoc.
DR Pfam; PF00008; EGF; 46.
DR Pfam; PF00683; TB; 9.
DR SMART; SM00179; EGF_CA; 42.
DR PROSITE; PS00010; ASX_HYDROXYL; 43.

```
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 38.
DR   PROSITE; PS01187; EGF_CA; 43.
KW   Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW   Repeat; Signal; Multigene family.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28   2871       FIBRILLIN 1.
FT   DOMAIN       81    112       EGF-LIKE 1.
FT   DOMAIN      115    146       EGF-LIKE 2.
FT   DOMAIN      147    178       EGF-LIKE 3.
FT   DOMAIN      246    287       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN      288    329       EGF-LIKE 5, CALCIUM-BINDING.
FT   DOMAIN      330    401       TGFBP 1.
FT   DOMAIN      402    446       PRO-RICH.
FT   DOMAIN      449    489       EGF-LIKE 6.
FT   DOMAIN      490    529       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN      530    571       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN      572    612       EGF-LIKE 9, CALCIUM-BINDING.
FT   DOMAIN      613    653       EGF-LIKE 10, CALCIUM-BINDING.
FT   DOMAIN      656    721       TGFBP 2.
FT   DOMAIN      723    764       EGF-LIKE 11, CALCIUM-BINDING.
FT   DOMAIN      765    806       EGF-LIKE 12, CALCIUM-BINDING.
FT   DOMAIN      807    846       EGF-LIKE 13, CALCIUM-BINDING.
FT   DOMAIN      910    951       EGF-LIKE 14, CALCIUM-BINDING.
FT   DOMAIN      952   1018       TGFBP 3.
FT   DOMAIN     1028   1069       EGF-LIKE 15, CALCIUM-BINDING.
FT   DOMAIN     1070   1112       EGF-LIKE 16, CALCIUM-BINDING.
FT   DOMAIN     1113   1154       EGF-LIKE 17, CALCIUM-BINDING.
FT   DOMAIN     1155   1196       EGF-LIKE 18, CALCIUM-BINDING.
FT   DOMAIN     1197   1237       EGF-LIKE 19, CALCIUM-BINDING.
FT   DOMAIN     1238   1279       EGF-LIKE 20, CALCIUM-BINDING.
FT   DOMAIN     1280   1321       EGF-LIKE 21, CALCIUM-BINDING.
FT   DOMAIN     1322   1362       EGF-LIKE 22, CALCIUM-BINDING.
FT   DOMAIN     1363   1403       EGF-LIKE 23, CALCIUM-BINDING.
FT   DOMAIN     1404   1445       EGF-LIKE 24, CALCIUM-BINDING.
FT   DOMAIN     1446   1486       EGF-LIKE 25, CALCIUM-BINDING.
FT   DOMAIN     1487   1527       EGF-LIKE 26, CALCIUM-BINDING.
FT   DOMAIN     1528   1599       TGFBP 4.
FT   DOMAIN     1606   1647       EGF-LIKE 27, CALCIUM-BINDING.
FT   DOMAIN     1648   1688       EGF-LIKE 28, CALCIUM-BINDING.
FT   DOMAIN     1689   1758       TGFBP 5.
FT   DOMAIN     1766   1807       EGF-LIKE 29, CALCIUM-BINDING.
FT   DOMAIN     1808   1848       EGF-LIKE 30, CALCIUM-BINDING.
FT   DOMAIN     1849   1890       EGF-LIKE 31, CALCIUM-BINDING.
FT   DOMAIN     1891   1929       EGF-LIKE 32, CALCIUM-BINDING.
FT   DOMAIN     1930   1972       EGF-LIKE 33, CALCIUM-BINDING.
FT   DOMAIN     1973   2012       EGF-LIKE 34, CALCIUM-BINDING.
FT   DOMAIN     2013   2054       EGF-LIKE 35, CALCIUM-BINDING.
FT   DOMAIN     2055   2121       TGFBP 6.
FT   DOMAIN     2127   2165       EGF-LIKE 36, CALCIUM-BINDING.
FT   DOMAIN     2166   2205       EGF-LIKE 37, CALCIUM-BINDING.
FT   DOMAIN     2206   2246       EGF-LIKE 38, CALCIUM-BINDING.
FT   DOMAIN     2247   2290       EGF-LIKE 39, CALCIUM-BINDING.
FT   DOMAIN     2291   2332       EGF-LIKE 40, CALCIUM-BINDING.
FT   DOMAIN     2333   2400       TGFBP 7.
FT   DOMAIN     2402   2443       EGF-LIKE 41, CALCIUM-BINDING.
FT   DOMAIN     2444   2484       EGF-LIKE 42, CALCIUM-BINDING.
FT   DOMAIN     2485   2523       EGF-LIKE 43, CALCIUM-BINDING.
FT   DOMAIN     2524   2566       EGF-LIKE 44, CALCIUM-BINDING.
FT   DOMAIN     2567   2606       EGF-LIKE 45, CALCIUM-BINDING.
FT   DOMAIN     2607   2647       EGF-LIKE 46, CALCIUM-BINDING.
FT   DOMAIN     2648   2687       EGF-LIKE 47, CALCIUM-BINDING.
FT   DISULFID     85     94       BY SIMILARITY.
FT   DISULFID     89    100       BY SIMILARITY.
FT   DISULFID    102    111       BY SIMILARITY.
FT   DISULFID    119    129       BY SIMILARITY.
FT   DISULFID    123    134       BY SIMILARITY.
FT   DISULFID    136    145       BY SIMILARITY.
FT   DISULFID    150    160       BY SIMILARITY.
FT   DISULFID    154    166       BY SIMILARITY.
FT   DISULFID    168    177       BY SIMILARITY.
FT   DISULFID    250    262       BY SIMILARITY.
FT   DISULFID    257    271       BY SIMILARITY.
FT   DISULFID    273    286       BY SIMILARITY.
FT   DISULFID    292    304       BY SIMILARITY.
FT   DISULFID    299    313       BY SIMILARITY.
FT   DISULFID    315    328       BY SIMILARITY.
FT   DISULFID    453    465       BY SIMILARITY.
FT   DISULFID    460    474       BY SIMILARITY.
FT   DISULFID    476    488       BY SIMILARITY.
FT   DISULFID    494    504       BY SIMILARITY.
FT   DISULFID    499    513       BY SIMILARITY.
FT   DISULFID    515    528       BY SIMILARITY.
FT   DISULFID    534    546       BY SIMILARITY.
FT   DISULFID    541    555       BY SIMILARITY.
FT   DISULFID    557    570       BY SIMILARITY.
FT   DISULFID    576    587       BY SIMILARITY.
FT   DISULFID    582    596       BY SIMILARITY.
FT   DISULFID    598    611       BY SIMILARITY.
FT   DISULFID    617    628       BY SIMILARITY.
FT   DISULFID    623    637       BY SIMILARITY.
FT   DISULFID    639    652       BY SIMILARITY.
FT   DISULFID    727    739       BY SIMILARITY.
FT   DISULFID    734    748       BY SIMILARITY.
FT   DISULFID    750    763       BY SIMILARITY.
FT   DISULFID    769    781       BY SIMILARITY.
FT   DISULFID    776    790       BY SIMILARITY.
FT   DISULFID    792    805       BY SIMILARITY.
FT   DISULFID    811    821       BY SIMILARITY.
FT   DISULFID    816    830       BY SIMILARITY.
FT   DISULFID    832    845       BY SIMILARITY.
FT   DISULFID    914    926       BY SIMILARITY.
FT   DISULFID    921    935       BY SIMILARITY.
FT   DISULFID    937    950       BY SIMILARITY.
FT   DISULFID   1032   1044       BY SIMILARITY.
FT   DISULFID   1039   1053       BY SIMILARITY.
FT   DISULFID   1055   1068       BY SIMILARITY.
FT   DISULFID   1074   1086       BY SIMILARITY.
FT   DISULFID   1081   1095       BY SIMILARITY.
FT   DISULFID   1097   1111       BY SIMILARITY.
FT   DISULFID   1117   1129       BY SIMILARITY.
FT   DISULFID   1124   1138       BY SIMILARITY.
FT   DISULFID   1140   1153       BY SIMILARITY.
FT   DISULFID   1159   1171       BY SIMILARITY.
FT   DISULFID   1166   1180       BY SIMILARITY.
FT   DISULFID   1182   1195       BY SIMILARITY.
FT   DISULFID   1201   1212       BY SIMILARITY.
FT   DISULFID   1208   1221       BY SIMILARITY.
FT   DISULFID   1223   1236       BY SIMILARITY.
FT   DISULFID   1242   1254       BY SIMILARITY.
FT   DISULFID   1249   1263       BY SIMILARITY.
FT   DISULFID   1265   1278       BY SIMILARITY.
FT   DISULFID   1284   1296       BY SIMILARITY.
FT   DISULFID   1291   1305       BY SIMILARITY.
FT   DISULFID   1307   1320       BY SIMILARITY.
FT   DISULFID   1326   1339       BY SIMILARITY.
FT   DISULFID   1333   1348       BY SIMILARITY.
FT   DISULFID   1350   1361       BY SIMILARITY.
FT   DISULFID   1367   1380       BY SIMILARITY.
FT   DISULFID   1374   1389       BY SIMILARITY.
FT   DISULFID   1391   1402       BY SIMILARITY.
FT   DISULFID   1408   1420       BY SIMILARITY.
FT   DISULFID   1415   1429       BY SIMILARITY.
FT   DISULFID   1431   1444       BY SIMILARITY.
FT   DISULFID   1450   1461       BY SIMILARITY.
FT   DISULFID   1456   1470       BY SIMILARITY.
FT   DISULFID   1472   1485       BY SIMILARITY.
FT   DISULFID   1491   1502       BY SIMILARITY.
FT   DISULFID   1497   1511       BY SIMILARITY.
FT   DISULFID   1513   1526       BY SIMILARITY.
FT   DISULFID   1610   1622       BY SIMILARITY.
FT   DISULFID   1617   1631       BY SIMILARITY.
FT   DISULFID   1633   1646       BY SIMILARITY.
FT   DISULFID   1652   1663       BY SIMILARITY.
FT   DISULFID   1658   1672       BY SIMILARITY.
FT   DISULFID   1674   1687       BY SIMILARITY.
```

```
  Query Match             7.3%;  Score 294.5;  DB 1;  Length 2871;
  Best Local Similarity   26.7%;  Pred. No. 1e-11;
  Matches   83;  Conservative   33;  Mismatches   94;  Indels  101;  Gaps   15;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  |:| | | |||: ||   || |  | ||
Db       1289 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCGRHAVCTNTA 1343

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
              | | | |  |:              ||  |||:|||   : :|     | || ||: | | |
Db       1344 GSFKCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKD 1393

Qy        155 GYLPRNGPEPFHPTTDATSCTEID--------CGTPP--EVPDGYIIGNYTSSLGSQVRY 204
              ||           | |  :||::|        ||       | ||             |
Db       1394 GY----------TGDGFTCTDLDECSENLNLCGNGQCLNAPGGY-------------RC 1429

Qy        205 ACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARY 264
              |  ||  ||     :|  :     | : |    | | |                  |: |
Db       1430 ECDMGF--VPSADGKACEDIDECSLPNI-CVFGTCHNLP---------------GLFRC 1470

Qy        265 VCQEGFE--SPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVR--------- 313
              |: |:|    ||                 ||::   ::  :  :  ||
Db       1471 ECEIGYELDRSGG---------------NCTDVNECLDPTTCISGNCVNTPGSYTCDC 1513

Qy        314 ---WQINSRRI 321
                 :::|  |:
Db       1514 SPDFELNPTRV 1524
```

```
RESULT 2
FBN1_HUMAN
ID   FBN1_HUMAN     STANDARD;      PRT;  2871 AA.
AC   P35555;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fibrillin 1 precursor.
GN   FBN1 OR FBN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=93372860; PubMed=8364578;
RA   Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA   Pangilinan T., Bonadio J.;
RT   "Genomic organization of the sequence coding for fibrillin, the
RT   defective gene product in Marfan syndrome.";
RL   Hum. Mol. Genet. 2:961-968(1993).
RN   [2]
RP   SEQUENCE OF 1-932 FROM N.A.
RC   TISSUE=Fibroblast, and Placenta;
RX   MEDLINE=94010947; PubMed=7691719;
RA   Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RT   "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT   multidomain structure and alternatively spliced exons at the 5'
RT   end.";
RL   Genomics 17:476-484(1993).
RN   [3]
RP   SEQUENCE OF 899-2871 FROM N.A.
RX   MEDLINE=91304568; PubMed=1852207;
RA   Maslen C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;
RT   "Partial sequence of a candidate gene for the Marfan syndrome.";
RL   Nature 352:334-337(1991).
RN   [4]
RP   SEQUENCE OF 813-1313 FROM N.A.
RX   MEDLINE=91304567; PubMed=1852206;
RA   Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA   Tsipouras P., Ramirez F., Hollister D.W.;
RT   "Linkage of Marfan syndrome and a phenotypically related disorder to
RT   two different fibrillin genes.";
RL   Nature 352:330-334(1991).
RN   [5]
RP   CHARACTERIZATION.
RX   MEDLINE=91317849; PubMed=1860873;
RA   Sakai L.Y., Keene D.R., Glanville R.W., Bachinger H.P.;
RT   "Purification and partial characterization of fibrillin, a cysteine-
RT   rich structural component of connective tissue microfibrils.";
RL   J. Biol. Chem. 266:14763-14770(1991).
RN   [6]
RP   STRUCTURE BY NMR OF 2054-2125.
RX   MEDLINE=98031893; PubMed=9362480;
RA   Yuan X., Downing A.K., Knott V., Handford P.A.;
RT   "Solution structure of the transforming growth factor beta-binding
RT   protein-like module, a domain associated with matrix fibrils.";
RL   EMBO J. 16:6659-6666(1997).
RN   [7]
RP   STRUCTURE BY NMR OF 2124-2205.
RX   MEDLINE=96144829; PubMed=8568869;
RA   Knott V., Downing A.K., Cardy C.M., Handford P.A.;
RT   "Calcium binding properties of an epidermal growth factor-like domain
RT   pair from human fibrillin-1.";
RL   J. Mol. Biol. 255:22-27(1996).
RN   [8]
RP   STRUCTURE BY NMR OF 2124-2205.
RX   MEDLINE=96222301; PubMed=8653794;
RA   Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,
RA   Handford P.A.;
RT   "Solution structure of a pair of calcium-binding epidermal growth
RT   factor-like domains: implications for the Marfan syndrome and other
RT   genetic disorders.";
RL   Cell 85:597-605(1996).
RN   [9]
RP   REVIEW ON MFS VARIANTS.
RX   MEDLINE=96174615; PubMed=8594563;
RA   Collod G., Beroud C., Soussi T., Junien C., Boileau C.;
RT   "Software and database for the analysis of mutations in the human
RT   FBN1 gene.";
RL   Nucleic Acids Res. 24:137-141(1996).
RN   [10]
RP   REVIEW ON MFS VARIANTS.
RX   MEDLINE=97169383; PubMed=9016526;
RA   Collod-Beroud G., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,
RA   Godfrey M., Hayward C., Karttunen L., Milewicz D., Peltonen L.,
RA   Richards R.I., Wang W., Junien C., Boileau C.;
RT   "Marfan Database (second edition): software and database for the
RT   analysis of mutations in the human FBN1 gene.";
RL   Nucleic Acids Res. 25:147-150(1997).
RN   [11]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=98062175; PubMed=9401003;
RA   Hayward C., Brock D.J.H.;
RT   "Fibrillin-1 mutations in Marfan syndrome and other type-1
RT   fibrillinopathies.";
RL   Hum. Mutat. 10:415-423(1997).
RN   [12]
RP   VARIANT MFS PRO-1137.
RX   MEDLINE=91304569; PubMed=1852208;
RA   Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,
RA   Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,
RA   Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;
RT   "Marfan syndrome caused by a recurrent de novo missense mutation in
RT   the fibrillin gene.";
RL   Nature 352:337-339(1991).
RN   [13]
RP   VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.
RX   MEDLINE=93250834; PubMed=1301946;
RA   Dietz H.C., Saraiva J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;
RT   "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome
RT   patients at cysteine residues in EGF-like domains.";
RL   Hum. Mutat. 1:366-374(1992).
RN   [14]
RP   VARIANT MFS SER-2307.
```

```
RX   MEDLINE=92235290; PubMed=1569206;
RA   Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr.,
RA   Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;
RT   "Marfan phenotype variability in a family segregating a missense
RT   mutation in the epidermal growth factor-like motif of the fibrillin
RT   gene.";
RL   J. Clin. Invest. 89:1674-1680(1992).
RN   [15]
RP   VARIANTS MFS ILE-548 AND ALA-723.
RX   MEDLINE=94010946; PubMed=8406497;
RA   Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,
RA   Pyeritz R.E., Francomano C.A.;
RT   "Four novel FBN1 mutations: significance for mutant transcript level
RT   and EGF-like domain calcium binding in the pathogenesis of Marfan
RT   syndrome.";
RL   Genomics 17:468-475(1993).
RN   [16]
RP   VARIANTS MFS SER-2144.
RX   MEDLINE=93278402; PubMed=8504310;
RA   Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;
RT   "A novel fibrillin mutation in the Marfan syndrome which could
RT   disrupt calcium binding of the epidermal growth factor-like module.";
RL   Hum. Mol. Genet. 2:475-477(1993).
RN   [17]
RP   VARIANTS MFS ARG-862; TYR-1117; PRO-1137 AND PHE-1589, AND VARIANT
RP   ALA-1148.
RX   MEDLINE=94108431; PubMed=8281141;
RA   Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,
RA   Berg M.A., Miller D.C., Francke U.;
RT   "Mutation screening of complete fibrillin-1 coding sequence: report
RT   of five new mutations, including two in 8-cysteine domains.";
RL   Hum. Mol. Genet. 2:1813-1821(1993).
RN   [18]
RP   VARIANTS MFS GLY-217 AMD ARG-2627.
RX   MEDLINE=95067970; PubMed=7977366;
RA   Karttunen L., Raghunath M., Loennqvist L., Peltonen L.;
RT   "A compound-heterozygous Marfan patient: two defective fibrillin
RT   alleles result in a lethal phenotype.";
RL   Am. J. Hum. Genet. 55:1083-1091(1994).
RN   [19]
RP   VARIANT EL LYS-2447.
RX   MEDLINE=94245249; PubMed=8188302;
RA   Lonnqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,
RA   Peltonen L.;
RT   "A novel mutation of the fibrillin gene causing ectopia lentis.";
RL   Genomics 19:573-576(1994).
RN   [20]
RP   VARIANT MFS CYS-627.
RX   MEDLINE=94272487; PubMed=8004112;
RA   Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;
RT   "Two novel mutations and a neutral polymorphism in EGF-like domains
RT   of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan
RT   syndrome patients.";
RL   Hum. Mol. Genet. 3:373-375(1994).
RN   [21]
RP   VARIANT MFS CYS-122.
RX   MEDLINE=94314977; PubMed=8040326;
RA   Stahl-Hallengren C., Ukkonen T., Kainulainen K., Kristofersson U.,
RA   Saxne T., Tornqvist K., Peltonen L.;
RT   "An extra cysteine in one of the non-calcium-binding epidermal growth
RT   factor-like motifs of the FBN1 polypeptide is connected to a novel
RT   variant of Marfan syndrome.";
RL   J. Clin. Invest. 94:709-713(1994).
RN   [22]
RP   VARIANT MFS TYR-1223.
RX   MEDLINE=94351682; PubMed=8071963;
RA   Hewett D.R., Lynch J.R., Child A., Sykes B.C.;
RT   "A new missense mutation of fibrillin in a patient with Marfan
RT   syndrome.";
RL   J. Med. Genet. 31:338-339(1994).
RN   [23]
RP   VARIANT MFS HIS-1170.
RX   MEDLINE=95174777; PubMed=7870075;
RA   Hayward C., Porteous M.E.M., Brock D.J.H.;
RT   "A novel mutation in the fibrillin gene (FBN1) in familial
RT   arachnodactyly.";
RL   Mol. Cell. Probes 8:325-327(1994).
RN   [24]
RP   VARIANTS MFS GLY-217; ASN-1023; ARG-1074; TYR-1242; ARG-1513;
RP   GLU-2127; TRP-2151; LYS-2447 AND ARG-2511.
RX   MEDLINE=94184368; PubMed=8136837;
RA   Kainulainen K., Karttunen L., Puhakka L., Sakai L., Peltonen L.;
RT   "Mutations in the fibrillin gene responsible for dominant ectopia
RT   lentis and neonatal Marfan syndrome.";
RL   Nat. Genet. 6:64-69(1994).
RN   [25]

  Query Match             7.2%;  Score 290.5;  DB 1;  Length 2871;
  Best Local Similarity   26.4%;  Pred. No. 1.9e-11;
  Matches   82;  Conservative   34;  Mismatches   94;  Indels  101;  Gaps   15;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  |:| | | |||: ||   || | | ||
Db       1289 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCGKHAVCTNTA 1343

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
              | | | | |:              || |||:|||   : :|     | || ||: | | :
Db       1344 GSFKCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKE 1393

Qy        155 GYLPRNGPEPFHPTTDATSCTEID--------CGTPP--EVPDGYIIGNYTSSLGSQVRY 204
              ||          | |  :||::|         ||      | ||              |
Db       1394 GY----------TGDGFTCTDLDECSENLNLCGNGQCLNAPGGY------------RC 1429

Qy        205 ACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARY 264
              |  ||  ||     :|  :     | : |    | | |                 |: |
Db       1430 ECDMGF--VPSADGKACEDIDECSLPNI-CVFGTCHNLP---------------GLFRC 1470

Qy        265 VCQEGFE--SPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVR--------- 313
              |: |:|    ||                    ||::   ::  :  :  ||
Db       1471 ECEIGYELDRSGG----------------NCTDVNECLDPTTCISGNCVNTPGSYICDC 1513

Qy        314 ---WQINSRRI 321
                 :::|  |:
Db       1514 PPDFELNPTRV 1524


RESULT 3
CD97_HUMAN
ID   CD97_HUMAN     STANDARD;      PRT;   835 AA.
AC   P48960;
DT   01-FEB-1996 (Rel. 33, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Leucocyte antigen CD97 precursor.
GN   CD97.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95363161; PubMed=7636245;
RA   Hamann J., Eichler W., Hamann D., Kerstens H.M.J., Poddighe P.J.,
RA   Hoovers J.M.N., Hartmann J.M., Strauss M., van Lier R.A.W.;
RT   "Expression cloning and chromosomal mapping of the leukocyte
RT   activation antigen CD97, a new seven-span transmembrane molecule of
RT   the secretion receptor superfamily with an unusual extracellular
RT   domain.";
RL   J. Immunol. 155:1942-1950(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Foreskin;
RX   MEDLINE=96230339; PubMed=8786105;
RA   Hamann J., Hartmann E., van Lier R.A.W.;
RT   "Structure of the human CD97 gene: exon shuffling has generated a new
RT   type of seven-span transmembrane molecule related to the secretin
```

```
RT   receptor superfamily.";
RL   Genomics 32:144-147(1996).
RN   [3]
RP   REVISIONS.
RA   Hamann J.;
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COULD BE A RECEPTOR POTENTIALLY INVOLVED IN BOTH
CC       ADHESION AND SIGNALING PROCESSES EARLY AFTER LEUKOCYTE ACTIVATION.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Contains 5 EGF-like domains.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   -!- DATABASE: NAME=PROW; NOTE=CD guide CD97 entry;
CC       WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd97.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X84700; CAA59173.1; -.
DR   EMBL; X94630; CAA64333.1; -.
DR   EMBL; X94631; CAA64333.1; JOINED.
DR   EMBL; X94632; CAA64333.1; JOINED.
DR   EMBL; X94633; CAA64333.1; JOINED.
DR   EMBL; Z99830; CAA64333.1; JOINED.
DR   EMBL; Z99831; CAA64333.1; JOINED.
DR   EMBL; X94634; CAA64333.1; JOINED.
DR   EMBL; X94635; CAA64333.1; JOINED.
DR   EMBL; X94636; CAA64333.1; JOINED.
DR   EMBL; X94637; CAA64333.1; JOINED.
DR   EMBL; X94638; CAA64333.1; JOINED.
DR   EMBL; X94639; CAA64333.1; JOINED.
DR   EMBL; X94640; CAA64333.1; JOINED.
DR   EMBL; X94641; CAA64333.1; JOINED.
DR   EMBL; X94642; CAA64333.1; JOINED.
DR   EMBL; X94643; CAA64333.1; JOINED.
DR   EMBL; X94644; CAA64333.1; JOINED.
DR   EMBL; X94645; CAA64333.1; JOINED.
DR   EMBL; X94646; CAA64333.1; JOINED.
DR   EMBL; X94647; CAA64333.1; JOINED.
DR   PIR; I37225; I37225.
DR   HSSP; P16109; 1FSB.
DR   Genew; HGNC:1711; CD97.
DR   MIM; 601211; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; TAS.
DR   GO; GO:0007155; P:cell adhesion; TAS.
DR   GO; GO:0006928; P:cell motility; TAS.
DR   GO; GO:0007267; P:cell-cell signaling; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   GO; GO:0006954; P:inflammatory response; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00179; EGF_CA; 4.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 4.
DR   PROSITE; PS01187; EGF_CA; 4.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   Cell adhesion; Receptor; G-protein coupled receptor; Repeat; Signal;
KW   Transmembrane; EGF-like domain; Glycoprotein.
```

```
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21    835       LEUCOCYTE ANTIGEN CD97.
FT   DOMAIN      21    552       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   553    572       POTENTIAL.
FT   DOMAIN     573    581       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   582    601       POTENTIAL.
FT   DOMAIN     602    620       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   621    642       POTENTIAL.
FT   DOMAIN     643    653       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   654    674       POTENTIAL.
FT   DOMAIN     675    691       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   692    712       POTENTIAL.
FT   DOMAIN     713    739       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   740    760       POTENTIAL.
FT   DOMAIN     761    766       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   767    789       POTENTIAL.
FT   DOMAIN     790    835       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      22     63       EGF-LIKE 1.
FT   DOMAIN      64    115       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     116    159       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     160    208       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     209    257       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     492    542       GPS.
FT   SITE       367    369       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    26     36       BY SIMILARITY.
FT   DISULFID    30     42       BY SIMILARITY.
FT   DISULFID    44     62       BY SIMILARITY.
FT   DISULFID    68     82       BY SIMILARITY.
FT   DISULFID    76     91       BY SIMILARITY.
FT   DISULFID    93    114       BY SIMILARITY.
FT   DISULFID   120    133       BY SIMILARITY.
FT   DISULFID   127    142       BY SIMILARITY.
FT   DISULFID   144    158       BY SIMILARITY.
FT   DISULFID   164    177       BY SIMILARITY.
FT   DISULFID   171    186       BY SIMILARITY.
FT   DISULFID   188    207       BY SIMILARITY.
FT   DISULFID   213    226       BY SIMILARITY.
FT   DISULFID   220    235       BY SIMILARITY.
FT   DISULFID   237    256       BY SIMILARITY.
FT   CARBOHYD    33     33       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   108    108       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   203    203       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   371    371       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   406    406       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   413    413       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   453    453       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   520    520       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   121    213       MISSING (IN REF. 1).
SQ   SEQUENCE   835 AA;  91941 MW;  050672E7A4C12A9E CRC64;
```

```
  Query Match             7.1%;  Score 287;  DB 1;  Length 835;
  Best Local Similarity   21.8%;  Pred. No. 7.1e-12;
  Matches  168;  Conservative   89;  Mismatches  242;  Indels  270;  Gaps   40;

Qy         29 GAPGPDGLDVCATCHEHATCQQREGKKICICNYGF------VGNGRTQCVDKNECQFGAT 82
              ||   |        | ::::|      | || ||      :        | | |||   :
Db         17 GAETQDSRGCARWCPQNSSCVNATA---CRCNPGFSSFSEIITTPTETCDDINECATPSK 73

Qy         83 LVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVS-GLCRHGGRC 141
              : ||  : | || | : |:|  ||   :  |||         | |:|||: :  ||:  | |
Db         74 VSCGKFSDCWNTEGSYDCVCSPGYEPVSGAKTFKNESENTCQDVDECQQNPRLCKSYGTC 133

Qy        142 VNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGS- 200
              ||| ||: | |: |:  :  ||    ||   ||    | |      : : |    :||
Db        134 VNTLGSYTCQCLPGF--KFIPEDPKVCTDVNECTS---GQNPCHSSTHCLNN----VGSY 184

Qy        201 QVRYACREGFFSV------PEDTVSSCTGLGTWESPKLHCQEIN-CGNPPEMRHAILVGN 253
              | |  || |:  :      | :||  |  :   | :  |    | |             ||:
Db        185 QCR--CRPGWQPIPGSPNGPNNTV--CEDVDECSSGQHQCDSSTVCFN--------TVGS 232

Qy        254 HSSRLGGVARYVCQEGFES----PGGKITSVCTEK--GTWRE----STLTCTEILTKIND 303
              :| |        |: |::     |  :  :|| :  . ||      : | :    |: |
```

```
Db        233 YSCR--------CRPGWKPRHGIPNNQKDTVCEDMTFSTWTPPPGVHSQTLSRFFDKVQD 284

Qy        304 VSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALY 363
              :   : |      :|  :   |  || :   ::     |
Db        285 LGRDSKT------SSAEVT--IQNVIKLVDMLMEAPGDVE---------------------- 316

Qy        364 PGTNYTVNISTAPPRRSMPAVIGFQTAEVDILEDDGSFNISIFNETCLKLNRRSRKVGSE 423
                    : ||| | : |     |  :  |||                ::::  :|   |
Db        317 ---------ALAPPVRHLIA-----TQLLSMLED------------IMRILAKSLPKGP- 349

Qy        424 HMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTV--------NVTLLR 475
                          |:: :  |                |: |:         |||: :
Db        350 ----------------FTYISPSN------ ----------TELTLMIQERGDKNVTMGQ 376

Qy        476 SPKRHSVQITIAT-----PPAVK--QTISNISG-FNETCLRWRSIKTADMEEMYLFHIWG 527
              |  |  :   :|       |||    :| |:      |   | | |::||:|
Db        377 SSARMKLNWAVAAGAEDPGPAVAGILSIQNMTTLLANASLNLHSKKQAELEEIY------ 430

Qy        528 QRWYQKEFAQEMTFNISSSSRDPEVCLDLRTGTNYNVSLRALSSELPVVISLTTQITEPP 587
                         || |                | || ||:   : :|
Db        431 ----------------ESSIR---------     -----GVQLRRLSAVNSIFLS--------- 451

Qy        588 LPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNAS 647
                        |             || | ||             |:          ||:
Db        452 ---------H---------NNTKELNSPI-  ------------LFA---------FSHLE 470

Qy        648 DADGYVAAELLAKDV-PDDAMEI------PTGDRLYYGEYYNAPL------KRGSDYCII 694
               :||    :  |||| |    |:          ||   | ::   :       | ||  |
Db        471 SSDGEAGRDPPAKDVMPGPRQELLCAFWKSDSDR---GGHWATEVCQVLGSKNGSTTCQC 527

Qy        695 LRITSEWNKVRRHSCAVWAQVKDSSLMLLQMAGVGLGSLAVVI-ILTFL 742
              ::| :  :  |       |:|  | |:   |: |    ::: |||||
Db        528 SHLSS-FTILMAH-----YDVEDWKLTLITRVGLALSLFCLLLCILTFL 570


RESULT 4
FBN1_BOVIN
ID   FBN1_BOVIN     STANDARD;      PRT;  2871 AA.
AC   P98133;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fibrillin 1 precursor (MP340).
GN   FBN1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=95137597; PubMed=7835900;
RA   Tilstra D.J., Potter K.A., Byers P.H.;
RT   "Sequence of the coding region of the bovine fibrillin cDNA and
RT   localization to bovine chromosome 10.";
RL   Genomics 23:480-485(1994).
RN   [2]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=96132851; PubMed=8557636;
RA   Gibson M.A., Hatzinikolas G., Kumaratilake J.S., Sandberg L.B.,
RA   Nicholl J.K., Sutherland G.R., Cleary E.G.;
RT   "Further characterization of proteins associated with elastic fiber
RT   microfibrils including the molecular cloning of MAGP-2 (MP25).";
RL   J. Biol. Chem. 271:1096-1103(1996).
CC   -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC       THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC       LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC   -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC       FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC       MICROFIBRILS.
CC   -!- SIMILARITY: Contains 47 EGF-like domains.
CC   -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L28748; AAA74122.1; -.
DR   PIR; A55567; A55567.
DR   HSSP; P35555; 1APJ.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   Pfam; PF00008; EGF; 46.
DR   Pfam; PF00683; TB; 9.
DR   SMART; SM00179; EGF_CA; 42.
DR   PROSITE; PS00010; ASX_HYDROXYL; 43.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 38.
DR   PROSITE; PS01187; EGF_CA; 43.
KW   Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW   Repeat; Signal; Multigene family.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28   2871       FIBRILLIN 1.
FT   DOMAIN       81    112       EGF-LIKE 1.
FT   DOMAIN      115    146       EGF-LIKE 2.
FT   DOMAIN      147    178       EGF-LIKE 3.
FT   DOMAIN      246    287       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN      288    329       EGF-LIKE 5, CALCIUM-BINDING.
FT   REPEAT      330    390       TGFBP 1.
FT   DOMAIN      392    446       PRO-RICH.
FT   DOMAIN      449    489       EGF-LIKE 6.
FT   DOMAIN      490    529       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN      530    571       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN      572    612       EGF-LIKE 9, CALCIUM-BINDING.
FT   DOMAIN      613    653       EGF-LIKE 10, CALCIUM-BINDING.
FT   REPEAT      654    722       TGFBP 2.
FT   DOMAIN      723    764       EGF-LIKE 11, CALCIUM-BINDING.
FT   DOMAIN      765    806       EGF-LIKE 12, CALCIUM-BINDING.
FT   DOMAIN      807    846       EGF-LIKE 13, CALCIUM-BINDING.
FT   DOMAIN      910    951       EGF-LIKE 14, CALCIUM-BINDING.
FT   REPEAT      952   1027       TGFBP 3.
FT   DOMAIN     1028   1069       EGF-LIKE 15, CALCIUM-BINDING.
FT   DOMAIN     1070   1112       EGF-LIKE 16, CALCIUM-BINDING.
FT   DOMAIN     1113   1154       EGF-LIKE 17, CALCIUM-BINDING.
FT   DOMAIN     1155   1196       EGF-LIKE 18, CALCIUM-BINDING.
FT   DOMAIN     1197   1237       EGF-LIKE 19, CALCIUM-BINDING.
FT   DOMAIN     1238   1279       EGF-LIKE 20, CALCIUM-BINDING.
FT   DOMAIN     1280   1321       EGF-LIKE 21, CALCIUM-BINDING.
FT   DOMAIN     1322   1362       EGF-LIKE 22, CALCIUM-BINDING.
FT   DOMAIN     1363   1403       EGF-LIKE 23, CALCIUM-BINDING.
FT   DOMAIN     1404   1445       EGF-LIKE 24, CALCIUM-BINDING.
FT   DOMAIN     1446   1486       EGF-LIKE 25, CALCIUM-BINDING.
FT   DOMAIN     1487   1527       EGF-LIKE 26, CALCIUM-BINDING.
FT   REPEAT     1528   1605       TGFBP 4.
FT   DOMAIN     1606   1647       EGF-LIKE 27, CALCIUM-BINDING.
FT   DOMAIN     1648   1688       EGF-LIKE 28, CALCIUM-BINDING.
FT   REPEAT     1689   1765       TGFBP 5.
FT   DOMAIN     1766   1807       EGF-LIKE 29, CALCIUM-BINDING.
FT   DOMAIN     1808   1848       EGF-LIKE 30, CALCIUM-BINDING.
FT   DOMAIN     1849   1890       EGF-LIKE 31, CALCIUM-BINDING.
FT   DOMAIN     1891   1929       EGF-LIKE 32, CALCIUM-BINDING.
FT   DOMAIN     1930   1972       EGF-LIKE 33, CALCIUM-BINDING.
FT   DOMAIN     1973   2012       EGF-LIKE 34, CALCIUM-BINDING.
FT   DOMAIN     2013   2054       EGF-LIKE 35, CALCIUM-BINDING.
FT   REPEAT     2055   2126       TGFBP 6.
FT   DOMAIN     2127   2165       EGF-LIKE 36, CALCIUM-BINDING.
FT   DOMAIN     2166   2205       EGF-LIKE 37, CALCIUM-BINDING.
FT   DOMAIN     2206   2246       EGF-LIKE 38, CALCIUM-BINDING.
FT   DOMAIN     2247   2290       EGF-LIKE 39, CALCIUM-BINDING.
```

```
FT   DOMAIN      2291    2332       EGF-LIKE 40, CALCIUM-BINDING.
FT   REPEAT      2333    2401       TGFBP 7.
FT   DOMAIN      2402    2443       EGF-LIKE 41, CALCIUM-BINDING.
FT   DOMAIN      2444    2484       EGF-LIKE 42, CALCIUM-BINDING.
FT   DOMAIN      2485    2523       EGF-LIKE 43, CALCIUM-BINDING.
FT   DOMAIN      2524    2566       EGF-LIKE 44, CALCIUM-BINDING.
FT   DOMAIN      2567    2606       EGF-LIKE 45, CALCIUM-BINDING.
FT   DOMAIN      2607    2647       EGF-LIKE 46, CALCIUM-BINDING.
FT   DOMAIN      2648    2687       EGF-LIKE 47, CALCIUM-BINDING.
FT   DISULFID      85      94       BY SIMILARITY.
FT   DISULFID      89     100       BY SIMILARITY.
FT   DISULFID     102     111       BY SIMILARITY.
FT   DISULFID     119     129       BY SIMILARITY.
FT   DISULFID     123     134       BY SIMILARITY.
FT   DISULFID     136     145       BY SIMILARITY.
FT   DISULFID     150     160       BY SIMILARITY.
FT   DISULFID     154     166       BY SIMILARITY.
FT   DISULFID     168     177       BY SIMILARITY.
FT   DISULFID     250     262       BY SIMILARITY.
FT   DISULFID     257     271       BY SIMILARITY.
FT   DISULFID     273     286       BY SIMILARITY.
FT   DISULFID     292     304       BY SIMILARITY.
FT   DISULFID     299     313       BY SIMILARITY.
FT   DISULFID     315     328       BY SIMILARITY.
FT   DISULFID     453     465       BY SIMILARITY.
FT   DISULFID     460     474       BY SIMILARITY.
FT   DISULFID     476     488       BY SIMILARITY.
FT   DISULFID     494     504       BY SIMILARITY.
FT   DISULFID     499     513       BY SIMILARITY.
FT   DISULFID     515     528       BY SIMILARITY.
FT   DISULFID     534     546       BY SIMILARITY.
FT   DISULFID     541     555       BY SIMILARITY.
FT   DISULFID     557     570       BY SIMILARITY.
FT   DISULFID     576     587       BY SIMILARITY.
FT   DISULFID     582     596       BY SIMILARITY.
FT   DISULFID     598     611       BY SIMILARITY.
FT   DISULFID     617     628       BY SIMILARITY.
FT   DISULFID     623     637       BY SIMILARITY.
FT   DISULFID     639     652       BY SIMILARITY.
FT   DISULFID     727     739       BY SIMILARITY.
FT   DISULFID     734     748       BY SIMILARITY.
FT   DISULFID     750     763       BY SIMILARITY.
FT   DISULFID     769     781       BY SIMILARITY.
FT   DISULFID     776     790       BY SIMILARITY.
FT   DISULFID     792     805       BY SIMILARITY.
FT   DISULFID     811     821       BY SIMILARITY.
FT   DISULFID     816     830       BY SIMILARITY.
FT   DISULFID     832     845       BY SIMILARITY.
FT   DISULFID     914     926       BY SIMILARITY.
FT   DISULFID     921     935       BY SIMILARITY.
FT   DISULFID     937     950       BY SIMILARITY.
FT   DISULFID    1032    1044       BY SIMILARITY.
FT   DISULFID    1039    1053       BY SIMILARITY.
FT   DISULFID    1055    1068       BY SIMILARITY.
FT   DISULFID    1074    1086       BY SIMILARITY.
FT   DISULFID    1081    1095       BY SIMILARITY.
FT   DISULFID    1097    1111       BY SIMILARITY.
FT   DISULFID    1117    1129       BY SIMILARITY.
FT   DISULFID    1124    1138       BY SIMILARITY.
FT   DISULFID    1140    1153       BY SIMILARITY.
FT   DISULFID    1159    1171       BY SIMILARITY.
FT   DISULFID    1166    1180       BY SIMILARITY.
FT   DISULFID    1182    1195       BY SIMILARITY.
FT   DISULFID    1201    1212       BY SIMILARITY.
FT   DISULFID    1208    1221       BY SIMILARITY.
FT   DISULFID    1223    1236       BY SIMILARITY.
FT   DISULFID    1242    1254       BY SIMILARITY.
FT   DISULFID    1249    1263       BY SIMILARITY.
FT   DISULFID    1265    1278       BY SIMILARITY.
FT   DISULFID    1284    1296       BY SIMILARITY.
FT   DISULFID    1291    1305       BY SIMILARITY.
FT   DISULFID    1307    1320       BY SIMILARITY.
FT   DISULFID    1326    1339       BY SIMILARITY.
FT   DISULFID    1333    1348       BY SIMILARITY.
FT   DISULFID    1350    1361       BY SIMILARITY.
FT   DISULFID    1367    1380       BY SIMILARITY.
FT   DISULFID    1374    1389       BY SIMILARITY.
FT   DISULFID    1391    1402       BY SIMILARITY.
FT   DISULFID    1408    1420       BY SIMILARITY.
FT   DISULFID    1415    1429       BY SIMILARITY.
FT   DISULFID    1431    1444       BY SIMILARITY.
FT   DISULFID    1450    1461       BY SIMILARITY.
FT   DISULFID    1456    1470       BY SIMILARITY.
FT   DISULFID    1472    1485       BY SIMILARITY.
FT   DISULFID    1491    1502       BY SIMILARITY.
FT   DISULFID    1497    1511       BY SIMILARITY.
FT   DISULFID    1513    1526       BY SIMILARITY.
FT   DISULFID    1610    1622       BY SIMILARITY.
FT   DISULFID    1617    1631       BY SIMILARITY.
FT   DISULFID    1633    1646       BY SIMILARITY.
FT   DISULFID    1652    1663       BY SIMILARITY.

  Query Match             7.0%;  Score 283.5;  DB 1;  Length 2871;
  Best Local Similarity  26.0%;  Pred. No. 5.8e-11;
  Matches   81;  Conservative   34;  Mismatches   95;  Indels  101;  Gaps   15;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  |:| | | |||: ||   |  |  | ||
Db       1289 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCDRHAVCTNTA 1343

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
              | | | |  |:           ||  |||:|||   : :|     | || ||: | | :
Db       1344 GSFKCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKE 1393

Qy        155 GYLPRNGPEPFHPTTDATSCTEID--------CGTPP--EVPDGYIIGNYTSSLGSQVRY 204
              ||           | |  :||::|        ||       | ||            |
Db       1394 GY----------TGDGFTCTDLDECSENLNLCGNGQCLNAPGGY------------RC 1429

Qy        205 ACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARY 264
              |  ||  ||     :|  :     | : |     | | |                 |: |
Db       1430 ECDMGF--VPSADGKACEDIDECSLPNI-CVFGTCHNLP---------------GLFRC 1470

Qy        265 VCQEGFE--SPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVR--------- 313
              |: |:|    ||                 ||::   ::  :  :  ||
Db       1471 ECEIGYELDRSGG----------------NCTDVNECLDPTTCISGNCVNTPGSYTCDC 1513

Qy        314 ---WQINSRRI 321
                 :::|  |:
Db       1514 PPDFELNPTRV 1524


RESULT 5
FBN2_MOUSE
ID   FBN2_MOUSE     STANDARD;      PRT;  2907 AA.
AC   Q61555; Q63957;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fibrillin 2 precursor.
GN   FBN2 OR FBN-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95263670; PubMed=7744963;
RA   Zhang H., Hu W., Ramirez F.;
RT   "Developmental expression of fibrillin genes suggests heterogeneity
RT   of extracellular microfibrils.";
RL   J. Cell Biol. 129:1165-1176(1995).
RN   [2]
RP   SEQUENCE OF 210-317 FROM N.A.
RX   MEDLINE=94140368; PubMed=8307578;
RA   Li X., Pereira L., Zhang H., Sanguineti C., Ramirez F., Bonadio J.,
RA   Francke U.;
```

```
RT   "Fibrillin genes map to regions of conserved mouse/human synteny on
RT   mouse chromosomes 2 and 18.";
RL   Genomics 18:667-672(1993).
CC   -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC       THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC       LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC   -!- SIMILARITY: Contains 47 EGF-like domains.
CC   -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L39790; AAA74908.1; -.
DR   EMBL; S69359; AAC60685.1; -.
DR   PIR; A57278; A57278.
DR   HSSP; P35555; 1EMN.
DR   MGD; MGI:95490; Fbn2.
DR   GO; GO:0030326; P:limb morphogenesis; IMP.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   Pfam; PF00008; EGF; 45.
DR   Pfam; PF00683; TB; 9.
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00179; EGF_CA; 43.
DR   PROSITE; PS00010; ASX_HYDROXYL; 43.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 36.
DR   PROSITE; PS01187; EGF_CA; 43.
KW   Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW   Repeat; Signal; Multigene family.
FT   SIGNAL       1     28       POTENTIAL.
FT   CHAIN       29   2907       FIBRILLIN 2.
FT   DOMAIN     111    142       EGF-LIKE 1
FT   DOMAIN     145    176       EGF-LIKE 2.
FT   DOMAIN     176    208       EGF-LIKE 3.
FT   DOMAIN     276    317       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN     318    359       EGF-LIKE 5, CALCIUM-BINDING.
FT   REPEAT     360    426       TGFBP 1.
FT   DOMAIN     487    527       EGF-LIKE 6.
FT   DOMAIN     528    567       EGF-LIKE 7  CALCIUM-BINDING.
FT   DOMAIN     568    609       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN     610    650       EGF-LIKE 9, CALCIUM-BINDING.
FT   DOMAIN     651    691       EGF-LIKE 10, CALCIUM-BINDING.
FT   REPEAT     692    760       TGFBP 2.
FT   DOMAIN     761    802       EGF-LIKE 11, CALCIUM-BINDING.
FT   DOMAIN     803    844       EGF-LIKE 12, CALCIUM-BINDING.
FT   DOMAIN     845    883       EGF-LIKE 13, CALCIUM-BINDING.
FT   DOMAIN     948    989       EGF-LIKE 14, CALCIUM-BINDING.
FT   REPEAT     990   1065       TGFBP 3.
FT   DOMAIN    1066   1107       EGF-LIKE 15, CALCIUM-BINDING.
FT   DOMAIN    1108   1150       EGF-LIKE 16, CALCIUM-BINDING.
FT   DOMAIN    1151   1192       EGF-LIKE 17, CALCIUM-BINDING.
FT   DOMAIN    1193   1234       EGF-LIKE 18, CALCIUM-BINDING.
FT   DOMAIN    1235   1275       EGF-LIKE 19, CALCIUM-BINDING.
FT   DOMAIN    1276   1317       EGF-LIKE 20, CALCIUM-BINDING.
FT   DOMAIN    1318   1359       EGF-LIKE 21, CALCIUM-BINDING.
FT   DOMAIN    1360   1400       EGF-LIKE 22, CALCIUM-BINDING.
FT   DOMAIN    1401   1441       EGF-LIKE 23, CALCIUM-BINDING.
FT   DOMAIN    1442   1483       EGF-LIKE 24, CALCIUM-BINDING.
FT   DOMAIN    1484   1524       EGF-LIKE 25, CALCIUM-BINDING.
FT   DOMAIN    1525   1565       EGF-LIKE 26, CALCIUM-BINDING.
FT   REPEAT    1566   1642       TGFBP 4.
FT   DOMAIN    1643   1684       EGF-LIKE 27, CALCIUM-BINDING.
FT   DOMAIN    1685   1726       EGF-LIKE 28, CALCIUM-BINDING.
FT   REPEAT    1727   1800       TGFBP 5.
FT   DOMAIN    1801   1842       EGF-LIKE 29, CALCIUM-BINDING.
FT   DOMAIN    1843   1884       EGF-LIKE 30, CALCIUM-BINDING.
FT   DOMAIN    1885   1926       EGF-LIKE 31, CALCIUM-BINDING.
FT   DOMAIN    1927   1965       EGF-LIKE 32, CALCIUM-BINDING.
FT   DOMAIN    1966   2008       EGF-LIKE 33, CALCIUM-BINDING.
FT   DOMAIN    2009   2048       EGF-LIKE 34, CALCIUM-BINDING.
FT   DOMAIN    2049   2090       EGF-LIKE 35, CALCIUM-BINDING.
FT   REPEAT    2091   2163       TGFBP 6.
FT   DOMAIN    2164   2205       EGF-LIKE 36, CALCIUM-BINDING.
FT   DOMAIN    2206   2245       EGF-LIKE 37, CALCIUM-BINDING.
FT   DOMAIN    2246   2286       EGF-LIKE 38, CALCIUM-BINDING.
FT   DOMAIN    2287   2330       EGF-LIKE 39, CALCIUM-BINDING.
FT   DOMAIN    2331   2372       EGF-LIKE 40, CALCIUM-BINDING.
FT   REPEAT    2373   2441       TGFBP 7.
FT   DOMAIN    2442   2483       EGF-LIKE 41, CALCIUM-BINDING.
FT   DOMAIN    2484   2524       EGF-LIKE 42, CALCIUM-BINDING.
FT   DOMAIN    2525   2563       EGF-LIKE 43, CALCIUM-BINDING.
FT   DOMAIN    2564   2606       EGF-LIKE 44, CALCIUM-BINDING.
FT   DOMAIN    2607   2646       EGF-LIKE 45, CALCIUM-BINDING.
FT   DOMAIN    2647   2687       EGF-LIKE 46, CALCIUM-BINDING.
FT   DOMAIN    2688   2727       EGF-LIKE 47, CALCIUM-BINDING.
FT   DISULFID   115    124       BY SIMILARITY.
FT   DISULFID   119    130       BY SIMILARITY.
FT   DISULFID   132    141       BY SIMILARITY.
FT   DISULFID   149    159       BY SIMILARITY.
FT   DISULFID   153    164       BY SIMILARITY.
FT   DISULFID   166    175       BY SIMILARITY.
FT   DISULFID   180    190       BY SIMILARITY.
FT   DISULFID   184    196       BY SIMILARITY.
FT   DISULFID   198    207       BY SIMILARITY.
FT   DISULFID   280    292       BY SIMILARITY.
FT   DISULFID   287    301       BY SIMILARITY.
FT   DISULFID   303    316       BY SIMILARITY.
FT   DISULFID   322    334       BY SIMILARITY.
FT   DISULFID   329    343       BY SIMILARITY.
FT   DISULFID   345    358       BY SIMILARITY.
FT   DISULFID   491    503       BY SIMILARITY.
FT   DISULFID   498    512       BY SIMILARITY.
FT   DISULFID   514    526       BY SIMILARITY.
FT   DISULFID   532    542       BY SIMILARITY.
FT   DISULFID   537    551       BY SIMILARITY.
FT   DISULFID   553    566       BY SIMILARITY.
FT   DISULFID   572    584       BY SIMILARITY.
FT   DISULFID   579    593       BY SIMILARITY.
FT   DISULFID   595    608       BY SIMILARITY.
FT   DISULFID   614    625       BY SIMILARITY.
FT   DISULFID   620    634       BY SIMILARITY.
FT   DISULFID   636    649       BY SIMILARITY.
FT   DISULFID   655    666       BY SIMILARITY.
FT   DISULFID   661    675       BY SIMILARITY.
FT   DISULFID   677    690       BY SIMILARITY.
FT   DISULFID   765    777       BY SIMILARITY.
FT   DISULFID   772    786       BY SIMILARITY.
FT   DISULFID   788    801       BY SIMILARITY.
FT   DISULFID   807    819       BY SIMILARITY.
FT   DISULFID   814    828       BY SIMILARITY.
FT   DISULFID   830    843       BY SIMILARITY.
FT   DISULFID   849    859       BY SIMILARITY.
FT   DISULFID   854    868       BY SIMILARITY.
FT   DISULFID   870    883       BY SIMILARITY.
FT   DISULFID   952    964       BY SIMILARITY.
FT   DISULFID   959    973       BY SIMILARITY.
FT   DISULFID   975    988       BY SIMILARITY.
FT   DISULFID  1070   1082       BY SIMILARITY.
FT   DISULFID  1077   1091       BY SIMILARITY.
FT   DISULFID  1093   1106       BY SIMILARITY.
FT   DISULFID  1112   1124       BY SIMILARITY.
FT   DISULFID  1119   1133       BY SIMILARITY.
FT   DISULFID  1135   1149       BY SIMILARITY.
FT   DISULFID  1155   1167       BY SIMILARITY.
FT   DISULFID  1162   1176       BY SIMILARITY.
FT   DISULFID  1178   1191       BY SIMILARITY.
FT   DISULFID  1197   1209       BY SIMILARITY.
```

```
FT   DISULFID   1204   1218       BY SIMILARITY.
FT   DISULFID   1220   1233       BY SIMILARITY.
FT   DISULFID   1239   1250       BY SIMILARITY.
FT   DISULFID   1246   1259       BY SIMILARITY.
FT   DISULFID   1261   1274       BY SIMILARITY.
FT   DISULFID   1280   1292       BY SIMILARITY.
FT   DISULFID   1287   1301       BY SIMILARITY.
FT   DISULFID   1303   1316       BY SIMILARITY.
FT   DISULFID   1322   1334       BY SIMILARITY.
FT   DISULFID   1329   1343       BY SIMILARITY.
FT   DISULFID   1345   1358       BY SIMILARITY.
FT   DISULFID   1364   1377       BY SIMILARITY.
FT   DISULFID   1371   1386       BY SIMILARITY.
FT   DISULFID   1388   1399       BY SIMILARITY.
FT   DISULFID   1405   1418       BY SIMILARITY.
FT   DISULFID   1412   1427       BY SIMILARITY.
FT   DISULFID   1429   1440       BY SIMILARITY.
FT   DISULFID   1446   1458       BY SIMILARITY.
FT   DISULFID   1453   1467       BY SIMILARITY.
FT   DISULFID   1469   1482       BY SIMILARITY.
FT   DISULFID   1488   1499       BY SIMILARITY.
FT   DISULFID   1494   1508       BY SIMILARITY.
FT   DISULFID   1510   1523       BY SIMILARITY.
FT   DISULFID   1529   1540       BY SIMILARITY.
FT   DISULFID   1535   1549       BY SIMILARITY.
FT   DISULFID   1551   1564       BY SIMILARITY.
FT   DISULFID   1647   1659       BY SIMILARITY.
FT   DISULFID   1654   1668       BY SIMILARITY.
FT   DISULFID   1670   1683       BY SIMILARITY.
FT   DISULFID   1689   1701       BY SIMILARITY.
FT   DISULFID   1696   1710       BY SIMILARITY.

  Query Match             7.0%;  Score 281;  DB 1;  Length 2907;
  Best Local Similarity   28.8%;  Pred. No. 8.7e-11;
  Matches   83;  Conservative   39;  Mismatches  114;  Indels   52;  Gaps   15;

Qy         25 RGAAGAPGPDGLDVCA-TCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGATL 83
              :|  |    |  ::  |  |  ||:|    |   | |  |:||||  :|:| :||  | |
Db       1353 KGTTGCTDVDECEIGAHNCDMHASCLNVPGSFKCSCREGWVGNG-IKCIDLDECANG-TH 1410

Qy         84 VCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCV 142
                |  :  | |||| : |  | ||:           ||  |:|:||| | : || : |:|:
Db       1411 QCSINAQCVNTPGSYRCACSEGF----------TGDGFTCSDVDECAENTNLCEN-GQCL 1459

Qy        143 NTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQ 201
              |  |:: | |  |          | | :|: || :|| |          :    ::|
Db       1460 NVPGAYRCECEMG---------FTPASDSRSCQDIDECSF-----QNICVFGTCNNLPGM 1505

Qy        202 VRYACREGFFSVPEDTVSSCTGLGTWESPKLNC-----------QEINCGNPPEMR-HAI 249
                | :|:    :  |  :|| :     | ::|            | ||  ||: : :|
Db       1506 FHCICDDGY--ELDRTGGNCTDIDECADP-INCVNGLCVNTPGRYECNC--PPDFQLNAT 1560

Qy        250 LVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEI 297
               ||   :|:|     |  |   |   | || |    :  |  :
Db       1561 GVGCVDNRVGN-----CYLKFGPRGDGSLSCNTEAGVGVSRSSCCCSL 1603


RESULT 6
FBN2_HUMAN
ID   FBN2_HUMAN     STANDARD;      PRT;  2911 AA.
AC   P35556;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Fibrillin 2 precursor.
GN   FBN2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94165150; PubMed=8120105;
RA   Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sanguineti C.,
RA   Bonadio J., Mecham R.P., Ramirez F.;
RT   "Structure and expression of fibrillin-2, a novel microfibrillar
RT   component preferentially located in elastic matrices.";
RL   J. Cell Biol. 124:855-863(1994).
RN   [2]
RP   SEQUENCE OF 752-1505 FROM N.A.
RX   MEDLINE=91304567; PubMed=1852206;
RA   Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA   Tsipouras P., Ramirez F., Hollister D.;
RT   "Linkage of Marfan syndrome and a phenotypically related disorder to
RT   two different fibrillin genes.";
RL   Nature 352:330-334(1991).
RN   [3]
RP   VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.
RX   MEDLINE=96083599; PubMed=7493032;
RA   Putnam E.A., Zhang H., Ramirez F., Milewicz D.M.;
RT   "Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,
RT   congenital contractural arachnodactyly.";
RL   Nat. Genet. 11:456-458(1995).
RN   [4]
RP   VARIANTS CCA HIS-1114.
RX   MEDLINE=98407789; PubMed=9737771;
RA   Babcock D., Gasner C., Francke U., Maslen C.;
RT   "A single mutation that results in an asp-to-his substitution and
RT   partial exon skipping in a family with congenital contractual
RT   arachnodactyly.";
RL   Hum. Genet. 103:22-28(1998).
RN   [5]
RP   VARIANTS CCA PHE-1141 AND TRP-1252.
RX   MEDLINE=20259236; PubMed=10797416;
RA   Belleh S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,
RA   Godfrey M.;
RT   "Two novel fibrillin-2 mutations in congenital contractural
RT   arachnodactyly.";
RL   Am. J. Med. Genet. 92:7-12(2000).
CC   -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC       THAT BINDS CALCIUM. FIBRILLIN-2-CONTAINING MICROFIBRILS REGULATE
CC       THE EARLY PROCESS OF ELASTIC FIBER ASSEMBLY.
CC   -!- DISEASE: DEFECTS IN FBN2 ARE THE CAUSE OF CONGENITAL CONTRACTURAL
CC       ARACHNODACTYLY (CCA) (ALSO KNOWN AS BEALS SYNDROME). CCA IS
CC       PHENOTYPICALLY SIMILAR TO MARFAN SYNDROME, BUT DOES NOT EFFECT THE
CC       AORTA AND THE EYES.
CC   -!- SIMILARITY: Contains 47 EGF-like domains.
CC   -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC   -!- DATABASE: NAME=Elastic Fiber Homepage; NOTE=Fibrillin 2 page;
CC       WWW="http://ef.wustl.edu/genes/FBN2.htm".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U03272; AAA18950.1; -.
DR   EMBL; X62009; -; NOT_ANNOTATED_CDS.
DR   PIR; A54105; A54105.
DR   HSSP; P35555; 1EMN.
DR   Genew; HGNC:3604; FBN2.
DR   MIM; 121050; -.
DR   GO; GO:0005578; C:extracellular matrix; TAS.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
DR   GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
DR   GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   Pfam; PF00008; EGF; 45.
DR   Pfam; PF00683; TB; 9.
```

```
DR   PRINTS; PR00010; EGFBLOOD.
DR   SMART; SM00179; EGF_CA; 43.
DR   PROSITE; PS00010; ASX_HYDROXYL; 43.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 37.
DR   PROSITE; PS01187; EGF_CA; 42.
KW   Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW   Repeat; Signal; Multigene family; Disease mutation; Polymorphism.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29   2911       FIBRILLIN 2.
FT   DOMAIN      111    142       EGF-LIKE 1.
FT   DOMAIN      145    176       EGF-LIKE 2.
FT   DOMAIN      176    207       EGF-LIKE 3.
FT   DOMAIN      275    316       EGF-LIKE 4, CALCIUM-BINDING.
FT   DOMAIN      317    358       EGF-LIKE 5, CALCIUM-BINDING.
FT   REPEAT      359    425       TGFBP 1.
FT   DOMAIN      493    533       EGF-LIKE 6
FT   DOMAIN      534    573       EGF-LIKE 7, CALCIUM-BINDING.
FT   DOMAIN      574    615       EGF-LIKE 8, CALCIUM-BINDING.
FT   DOMAIN      616    656       EGF-LIKE 9, CALCIUM-BINDING.
FT   DOMAIN      657    697       EGF-LIKE 10, CALCIUM-BINDING.
FT   REPEAT      698    766       TGFBP 2.
FT   DOMAIN      767    808       EGF-LIKE 11, CALCIUM-BINDING.
FT   DOMAIN      809    850       EGF-LIKE 12, CALCIUM-BINDING.
FT   DOMAIN      851    890       EGF-LIKE 13, CALCIUM-BINDING.
FT   DOMAIN      954    995       EGF-LIKE 14, CALCIUM-BINDING.
FT   REPEAT      996   1071       TGFBP 3.
FT   DOMAIN     1072   1113       EGF-LIKE 15, CALCIUM-BINDING.
FT   DOMAIN     1114   1156       EGF-LIKE 16, CALCIUM-BINDING.
FT   DOMAIN     1157   1198       EGF-LIKE 17, CALCIUM-BINDING.
FT   DOMAIN     1199   1240       EGF-LIKE 18, CALCIUM-BINDING.
FT   DOMAIN     1241   1281       EGF-LIKE 19, CALCIUM-BINDING.
FT   DOMAIN     1282   1323       EGF-LIKE 20, CALCIUM-BINDING.
FT   DOMAIN     1324   1365       EGF-LIKE 21, CALCIUM-BINDING.
FT   DOMAIN     1366   1406       EGF-LIKE 22, CALCIUM-BINDING.
FT   DOMAIN     1407   1447       EGF-LIKE 23, CALCIUM-BINDING.
FT   DOMAIN     1448   1489       EGF-LIKE 24, CALCIUM-BINDING.
FT   DOMAIN     1490   1530       EGF-LIKE 25, CALCIUM-BINDING.
FT   DOMAIN     1531   1571       EGF-LIKE 26, CALCIUM-BINDING.
FT   REPEAT     1572   1648       TGFBP 4.
FT   DOMAIN     1649   1690       EGF-LIKE 27, CALCIUM-BINDING.
FT   DOMAIN     1691   1732       EGF-LIKE 28, CALCIUM-BINDING.
FT   REPEAT     1733   1806       TGFBP 5.
FT   DOMAIN     1807   1848       EGF-LIKE 29, CALCIUM-BINDING.
FT   DOMAIN     1849   1890       EGF-LIKE 30, CALCIUM-BINDING.
FT   DOMAIN     1891   1932       EGF-LIKE 31, CALCIUM-BINDING.
FT   DOMAIN     1933   1971       EGF-LIKE 32, CALCIUM-BINDING.
FT   DOMAIN     1972   2014       EGF-LIKE 33, CALCIUM-BINDING.
FT   DOMAIN     2015   2054       EGF-LIKE 34, CALCIUM-BINDING.
FT   DOMAIN     2055   2096       EGF-LIKE 35, CALCIUM-BINDING.
FT   REPEAT     2097   2169       TGFBP 6.
FT   DOMAIN     2170   2211       EGF-LIKE 36, CALCIUM-BINDING.
FT   DOMAIN     2212   2251       EGF-LIKE 37, CALCIUM-BINDING.
FT   DOMAIN     2252   2292       EGF-LIKE 38, CALCIUM-BINDING.
FT   DOMAIN     2293   2336       EGF-LIKE 39, CALCIUM-BINDING.
FT   DOMAIN     2337   2378       EGF-LIKE 40, CALCIUM-BINDING.
FT   REPEAT     2379   2447       TGFBP 7.
FT   DOMAIN     2448   2489       EGF-LIKE 41, CALCIUM-BINDING.
FT   DOMAIN     2490   2530       EGF-LIKE 42, CALCIUM-BINDING.
FT   DOMAIN     2531   2569       EGF-LIKE 43, CALCIUM-BINDING.
FT   DOMAIN     2570   2612       EGF-LIKE 44, CALCIUM-BINDING.
FT   DOMAIN     2613   2652       EGF-LIKE 45, CALCIUM-BINDING.
FT   DOMAIN     2653   2693       EGF-LIKE 46, CALCIUM-BINDING.
FT   DOMAIN     2694   2733       EGF-LIKE 47, CALCIUM-BINDING.
FT   DISULFID    115    124       BY SIMILARITY.
FT   DISULFID    119    130       BY SIMILARITY.
FT   DISULFID    132    141       BY SIMILARITY.
FT   DISULFID    149    159       BY SIMILARITY.
FT   DISULFID    153    164       BY SIMILARITY.
FT   DISULFID    166    175       BY SIMILARITY.
FT   DISULFID    180    190       BY SIMILARITY.
FT   DISULFID    184    195       BY SIMILARITY.
FT   DISULFID    197    206       BY SIMILARITY.
FT   DISULFID    279    291       BY SIMILARITY.
FT   DISULFID    286    300       BY SIMILARITY.
FT   DISULFID    302    315       BY SIMILARITY.
FT   DISULFID    321    333       BY SIMILARITY.
FT   DISULFID    328    342       BY SIMILARITY.
FT   DISULFID    344    357       BY SIMILARITY.
FT   DISULFID    497    509       BY SIMILARITY.
FT   DISULFID    504    518       BY SIMILARITY.
FT   DISULFID    520    532       BY SIMILARITY.
FT   DISULFID    538    548       BY SIMILARITY.
FT   DISULFID    543    557       BY SIMILARITY.
FT   DISULFID    559    572       BY SIMILARITY.
FT   DISULFID    578    590       BY SIMILARITY.
FT   DISULFID    585    599       BY SIMILARITY.
FT   DISULFID    601    614       BY SIMILARITY.
FT   DISULFID    620    631       BY SIMILARITY.
FT   DISULFID    626    640       BY SIMILARITY.
FT   DISULFID    642    655       BY SIMILARITY.
FT   DISULFID    661    672       BY SIMILARITY.
FT   DISULFID    667    681       BY SIMILARITY.
FT   DISULFID    683    696       BY SIMILARITY.
FT   DISULFID    771    783       BY SIMILARITY.
FT   DISULFID    778    792       BY SIMILARITY.
FT   DISULFID    794    807       BY SIMILARITY.
FT   DISULFID    813    825       BY SIMILARITY.
FT   DISULFID    820    834       BY SIMILARITY.
FT   DISULFID    836    849       BY SIMILARITY.
FT   DISULFID    855    865       BY SIMILARITY.
FT   DISULFID    860    874       BY SIMILARITY.
FT   DISULFID    876    889       BY SIMILARITY.
FT   DISULFID    958    970       BY SIMILARITY.
FT   DISULFID    965    979       BY SIMILARITY.
FT   DISULFID    981    994       BY SIMILARITY.
FT   DISULFID   1076   1088       BY SIMILARITY.
FT   DISULFID   1083   1097       BY SIMILARITY.
FT   DISULFID   1099   1112       BY SIMILARITY.
FT   DISULFID   1118   1130       BY SIMILARITY.
FT   DISULFID   1125   1139       BY SIMILARITY.
FT   DISULFID   1141   1155       BY SIMILARITY.
FT   DISULFID   1161   1173       BY SIMILARITY.

  Query Match              6.9%;  Score 278.5;  DB 1;  Length 2911;
  Best Local Similarity   28.8%;  Pred. No. 1.3e-10;
  Matches   90;  Conservative   30;  Mismatches  126;  Indels   67;  Gaps   17;

Qy          2 GRGPWDAGPSRRLLPLLLLLGLARG------AAGAPGP--DGLDV----CATCHEHAT-- 47
              |  ||   |     :|:    | : |        ||  ||   || :        |  ||
Db        449 GYGPGGTG----FIPIPGGNGFSPGVGGAGVGAGGQGPIITGLTILNQTIDICKHHANLC 504

Qy         48 ----CQQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICL 103
                   |      | ||  |:  :    |:| :||   :  | |   | |||| :|| |
Db        505 LNGRCIPTVSSYRCECNMGYKQDANGDCIDVDEC---TSNPCTN-GDCVNTPGSYYCKCH 560

Qy        104 EGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPE 163
                |:: |   :          | |||||  :|:    |||||: |||:| |  |
Db        561 AGFQRTPTKQA--------CIDIDECIQNGVLCKNGRCVNSDGSFQCICNAG-------- 604

Qy        164 PFHPTTDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCT 222
                |  |||  :| : | | |      :|  |            :  |: ||   |      ||
Db        605 -FELTTDGKNCVDHDECTTTNMCLNGMCINE-----DGSFKCICKPGFVLAPNGRY--CT 656

Qy        223 GLGTWESPKL----HC----QEINCGNPP----EMRHAILVGNH--SSRLGGVARYVCQE 268
                :   ::| :    ||      |  ||     |  : |  |  |:  ||: : ||
Db        657 DVDECQTPGICMNGHCINSEGSFRCDCPPGLAVGMDGRVCVDTHMRSTCYGGIKKGVCVR 716

Qy        269 GFESPGGKITSVC 281
              |  ||    | |
Db        717 PF--PGAVTKSEC 727


RESULT 7
UROM_HUMAN
ID   UROM_HUMAN     STANDARD;      PRT;   640 AA.
```

```
AC   P07911;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Uromodulin precursor (Tamm-Horsfall urinary glycoprotein) (THP).
GN   UMOD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87177970; PubMed=3453112;
RA   Pennica D., Kohr W.J., Kuang W.-J., Glaister D., Aggarwal B.B.,
RA   Chen E.Y., Goeddel D.V.;
RT   "Identification of human uromodulin as the Tamm-Horsfall urinary
RT   glycoprotein.";
RL   Science 236:83-88(1987).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=87319675; PubMed=3498215;
RA   Hession C., Decker J.M., Sherblom A.P., Kumar S., Yue C.C.,
RA   Mattaliano R.J., Tizard R., Kawashima E., Schmeissner U.,
RA   Heletky S., Chow E.P., Burne C.A., Shaw A., Muchmore A.V.;
RT   "Uromodulin (Tamm-Horsfall glycoprotein): a renal ligand for
RT   lymphokines.";
RL   Science 237:1479-1484(1987).
RN   [3]
RP   GPI-ANCHOR.
RX   MEDLINE=91065873; PubMed=2249987;
RA   Rindler M.J., Naik S.S., Li N., Hoops T.C., Peraldi M.-N.;
RT   "Uromodulin (Tamm-Horsfall glycoprotein/uromucoid) is a
RT   phosphatidylinositol-linked membrane protein.";
RL   J. Biol. Chem. 265:20784-20789(1990).
CC   -!- FUNCTION: NOT KNOWN. MAY PLAY A ROLE IN REGULATING THE CIRCULATING
CC       ACTIVITY OF CYTOKINES AS IT BINDS TO IL-1, IL-2 AND TNF WITH HIGH
CC       AFFINITY.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR,
CC       THEN CLEAVED TO PRODUCE A SOLUBLE FORM WHICH IS SECRETED IN
CC       URINE.
CC   -!- TISSUE SPECIFICITY: SYNTHESIZED BY THE KIDNEYS AND IS THE MOST
CC       ABUNDANT PROTEIN IN NORMAL HUMAN URINE.
CC   -!- SIMILARITY: Contains 3 EGF-like domains.
CC   -!- SIMILARITY: Contains 1 ZP domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15881; AAA36798.1; -.
DR   EMBL; M17778; AAA36799.1; -.
DR   PIR; A30452; A30452.
DR   HSSP; P07204; 1ADX.
DR   GlycoSuiteDB; P07911; -.
DR   Genew; HGNC:12559; UMOD.
DR   MIM; 191845; -.
DR   GO; GO:0015025; F:GPI-anchored membrane-bound receptor; TAS.
DR   GO; GO:0006968; P:cellular defense response; TAS.
DR   GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR001507; Endoglin/CD105.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF00100; zona_pellucida; 1.
DR   PRINTS; PR00023; ZPELLUCIDA.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00241; ZP; 1.
DR   PROSITE; PS00682; ZP_DOMAIN; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 2.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01187; EGF_CA; 2.
KW   Glycoprotein; Signal; Membrane; GPI-anchor; EGF-like domain.
FT   SIGNAL        1    24
FT   CHAIN        25   640       UROMODULIN.
FT   DOMAIN       28    64       EGF-LIKE 1.
FT   DOMAIN       65   107       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      108   149       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      334   585       ZP.
FT   DISULFID     32    41       BY SIMILARITY.
FT   DISULFID     35    50       BY SIMILARITY.
FT   DISULFID     52    63       BY SIMILARITY.
FT   DISULFID     69    83       BY SIMILARITY.
FT   DISULFID     77    92       BY SIMILARITY.
FT   DISULFID     94   106       BY SIMILARITY.
FT   DISULFID    112   126       BY SIMILARITY.
FT   DISULFID    120   135       BY SIMILARITY.
FT   DISULFID    137   148       BY SIMILARITY.
FT   CARBOHYD     38    38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     76    76       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     80    80       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    232   232       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    275   275       N-LINKED (GLCNAC. . .).
FT                               /FTId=CAR_000178.
FT   CARBOHYD    322   322       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    396   396       N-LINKED (GLCNAC. . .).
FT   CONFLICT    565   565       H -> D (IN REF. 2).
SQ   SEQUENCE   640 AA;  69760 MW;  D26A07A76353AE48 CRC64;
```

```
  Query Match             6.9%;  Score 277.5;  DB 1;  Length 640;
  Best Local Similarity   23.4%;  Pred. No. 2.2e-11;
  Matches  163;  Conservative   77;  Mismatches  217;  Indels  239;  Gaps   38;

Qy         39 CATCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGF 98
              |: || :||| : |    | |  || |:| | ||| :||       |  ::|| |||| |
Db         32 CSECHSNATCTEDEAVTTCTCQEGFTGDGLT-CVDLDECAIPGAHNCSANSSCVNTPGSF 90

Qy         99 YCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGL--CRHGGRCVNTHGSFECYCMDGY 156
              |:| ||:|         : |  |||:|||   ||  |    |||  ||: | |  ||
Db         91 SCVCPEGFRL---------SPGLGCTDVDECAEPGLSHCHALATCVNVVGSYLCVCPAGY 141

Qy        157 --------------------LPRN------GPEPFHPTTD---------------------- 170
                                  :|        |   | | |
Db        142 RGDGWHCECSPGSCGPGLDCVPEGDALVCADPCQAHRTLDEYWRSTEYGEGYACDTDLRG 201

Qy        171 ------------ATSCTEI-DC---------GTPPEVPDGYIIGNYTS---------SLG 199
                       | :|  :  |           || |   :| :    :
Db        202 WYRFVGQGGARMAETCVPVLRCNTAAPMWLNGTHPSSDEGIVSRKACAHWSGHCCLWDAS 261

Qy        200 SQVRYACREGFF-----SVPEDTVSSCTG----LGTWE----------SPKLHC---QE 236
              ||: ||  |::     : ||  :: ||       || |              : : ||  |:
Db        262 VQVK-ACAGGYYVYNLTAPPECHLAYCTDPSSVEGTCEECSIDEDCKSNNGRWHCQCKQD 320

Qy        237 INCGNPPEMRHAILVGNHSSRLGGVARYVCQ---EGFESPGGKI-----TSVCT-----E 283
              |  :  : | :  | :  :   |:   ||    ||:    |:       | |:     :
Db        321 FNITDISLLEHRLECGANDMK---VSLGKCQLKSLGPD----KVFMYLSDSRCSGFNDRD 373

Qy        284 KGTWRESTLT------CTEILTKINDVSLFNDTC-VRWQINSRRINPKISYV-------- 328
              |  | :|       |  :||:    : :::|  :  :|  | :| ||::
Db        374 NRDW-VSVVTPARDGPCGTVLTRNETHATYSNTLYLADEIIIRDLNIKINFACSYPLDMK 432

Qy        329 ISIKGQRLDPMES---VREETVNLTTDSRTPEVCLALYPGTNYTVNISTAPPRRSMPAVI 385
              :|:|   | || |   :|    :  |      | :||:   :||              |
Db        433 VSLK-TALQPMVSALNIRVGGTGMFT------VRMALFQTPSYT----------QP--- 471

Qy        386 GFQTAEVDLLEDDGSFNISIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATS 445
              :| : | |      |    ::  | |      ||          | :|    |    |:||
Db        472 -YQGSSVTL-----STEAFLYVGTMLDGGDLSR---------PALLMTNCYATPSSNATD 516

Qy        446 --FNFTTREQVPVVCLDLYPTTDYTVNVTL--LRSPKRHSVQITIATPPAVKQTISNISG 501
                  |  ::: |        | | |: |      |  | |||:
```

```
Db        517 PLKYFIIQDRCP-------HTRDSTIQVVENGESSQGRFSVQM---------------- 552

Qy        502 FNETCLRWRSIKTADME----EMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLR 557
                  :|     |:     |:||     :   |     |  | | |  |     |:|
Db        553 -------FRFAGNYDLVYLHCEVYLCDTMNEKC--KPTCSGTRFR-SGSVIDQSRVLNLG 602

Qy        558 PGTNYNV---------SLRALSSELPVVISLTTQIT 584
              | | |          ||  |   ||:::| |  :|
Db        603 PITRKGVQATVSRAFSSLGLLKVWLPLLLSATLTLT 638
```

```
RESULT 8
EMR1_HUMAN
ID   EMR1_HUMAN     STANDARD;      PRT;   886 AA.
AC   Q14246;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Cell surface glycoprotein EMR1 precursor (EMR1 hormone receptor).
GN   EMR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95324926; PubMed=7601460;
RA   Baud V., Chissoe S.L., Viegas-Pequignot E., Diriong S., N'Guyen V.C.,
RA   Roe B.A., Lipinski M.;
RT   "EMR1, an unusual member in the family of hormone receptors with
RT   seven transmembrane segments.";
RL   Genomics 26:334-344(1995).
CC   -!- FUNCTION: PROBABLY INVOLVED IN CELLULAR RESPONSE TO A HORMONE OR
CC       AN INTERACTION WITH A PROTEIN LIGAND.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: WIDE EXPRESSION; INCREASED LEVELS IN
CC       PERIPHERAL BLOOD MONONUCLEAR CELLS.
CC   -!- PTM: N- AND O-GLYCOSYLATED; (POSSIBLE).
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X81479; CAA57232.1; -.
DR   PIR; A57172; A57172.
DR   HSSP; P00736; 1APQ.
DR   Genew; HGNC:3336; EMR1.
DR   MIM; 600493; -.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0004930; F:G-protein coupled receptor activity; TAS.
DR   GO; GO:0007155; P:cell adhesion; TAS.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00179; EGF_CA; 5.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 6.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   G-protein coupled receptor; Transmembrane; Receptor; Glycoprotein;
KW   EGF-like domain; Repeat; Signal.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18    886       CELL SURFACE GLYCOPROTEIN EMR1.
FT   DOMAIN       18    599       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    600    627       POTENTIAL.
FT   DOMAIN      628    634       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    635    656       POTENTIAL.
FT   DOMAIN      657    666       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    667    690       POTENTIAL.
FT   DOMAIN      691    709       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    710    731       POTENTIAL.
FT   DOMAIN      732    747       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    748    776       POTENTIAL.
FT   DOMAIN      777    794       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    795    814       POTENTIAL.
FT   DOMAIN      815    829       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    830    852       POTENTIAL.
FT   DOMAIN      853    886       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31     79       EGF-LIKE 1.
FT   DOMAIN       80    131       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      132    171       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      172    213       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      221    267       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      268    316       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      547    596       GPS.
FT   DOMAIN      317    599       SER/THR-RICH.
FT   DISULFID     35     47       BY SIMILARITY.
FT   DISULFID     41     56       BY SIMILARITY.
FT   DISULFID     58     78       BY SIMILARITY.
FT   DISULFID     84     97       BY SIMILARITY.
FT   DISULFID     91    106       BY SIMILARITY.
FT   DISULFID    108    130       BY SIMILARITY.
FT   DISULFID    136    148       BY SIMILARITY.
FT   DISULFID    142    157       BY SIMILARITY.
FT   DISULFID    159    170       BY SIMILARITY.
FT   DISULFID    176    188       BY SIMILARITY.
FT   DISULFID    182    197       BY SIMILARITY.
FT   DISULFID    199    212       BY SIMILARITY.
FT   DISULFID    225    235       BY SIMILARITY.
FT   DISULFID    229    244       BY SIMILARITY.
FT   DISULFID    246    266       BY SIMILARITY.
FT   DISULFID    272    285       BY SIMILARITY.
FT   DISULFID    279    294       BY SIMILARITY.
FT   DISULFID    296    315       BY SIMILARITY.
FT   CARBOHYD     94     94       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     99     99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    127    127       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    167    167       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    189    189       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    194    194       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    232    232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    258    258       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    312    312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    366    366       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    375    375       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    448    448       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    661    661       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   886 AA;  97680 MW;  7456CA56FB624D99 CRC64;
```

```
  Query Match             6.8%;  Score 274.5;  DB 1;  Length 886;
  Best Local Similarity   22.8%;  Pred. No. 5.3e-11;
  Matches   89;  Conservative   53;  Mismatches  125;  Indels  123;  Gaps   17;

Qy         42 CHEHATCQQREGKKICICNYGFV-GNGR-------TQCVDKNECQFGATLVCGNHTSCHN 93
                |  :|||         | |  ||:  ||:        :| | :||   :   || ::|| |
Db         41 CPAYATCTNTVDSYYCTCKQGFLSSNGQNHFKDPGVRCKDIDECS-QSPQPCGPNSSCKN 99

Qy         94 TPGGFYCICLEGYRATNNNKTFIPNDGTF-CTDIDECEVSGLCRHGGRCVNTHGSFECYC 152
                 | : | ||:|: :   |        | | ||||:||  | :|     |||: ||: | |
Db        100 LSGRYKCSCLDGFSSPTGNDWVPGKPGNFSCTDINECLTSRVCPEHSDCVNSMGSYSCSC 159
```

```
Qy        153 MDGYLPRNGPEPFHPTTDATSCTEI-DCGTPPEVPD----GYIIGNYTSSLGSQVRYACR 207
Db        160 QVGFISRN-----------STCEDVNECADPRACPEHATCNNTVGNYSC--------FCN 200

Qy        208 EGFFSVPEDTVSSCTGLGTWESPKLHCQEIN-----------CGNPPEMRHAILVGNHSS 256
Db        201 PGFES--SSGHLSCQGL------KASCEDIDECTEMCPINSTCTNTP------------- 239

Qy        257 RLGGVARYVCQEGFESPGGKI-------------------------TSVCTE--------- 283
Db        240 ---GSYFCTCHPGFAPSSGQLNFTDQGVECRDIDECRQDPSTCGPNSICTNALGSYSCGC 296

Qy        284 --------KGTWRESTLTCTEILTKINDVSLFNDTCVRWQINSRRINP-------KISYV 328
Db        297 IVGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNKQIQQCQEGTAVKPAYVSFCAQINNI 356

Qy        329 ISIKGQRLDPMESVREETVNL--TTDSRTP 356
Db        357 FSV----LDKVCENKTTVVSLKNTTESFVP 382
```

RESULT 9
FBN1_PIG
ID FBN1_PIG STANDARD; PRT; 2871 AA.
AC Q9TV36;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99156858; PubMed=10036187;
RA Biery N.J., Eldadah Z.A., Moore C.S., Stetten G., Spencer F.,
RA Dietz H.C.;
RT "Revised genomic organization of FBN1 and significance for regulated
RT gene expression.";
RL Genomics 56:70-77(1999).
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS.
CC -!- SIMILARITY: Contains 47 EGF-like domains.
CC -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; AF073800; AAD50328.1; -.
DR HSSP; P35555; 1APJ.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002212; Fibril-assoc.
DR Pfam; PF00008; EGF; 45.
DR Pfam; PF00683; TB; 9.
DR SMART; SM00179; EGF_CA; 40.
DR PROSITE; PS00010; ASX_HYDROXYL; 41.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 36.
DR PROSITE; PS01187; EGF_CA; 41.
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.

FT	SIGNAL	1	27	POTENTIAL.
FT	CHAIN	28	2871	FIBRILLIN 1.
FT	DOMAIN	81	112	EGF-LIKE 1.
FT	DOMAIN	115	146	EGF-LIKE 2.
FT	DOMAIN	147	178	EGF-LIKE 3.
FT	DOMAIN	246	287	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	288	329	EGF-LIKE 5, CALCIUM-BINDING.
FT	REPEAT	330	390	TGFBP 1.
FT	DOMAIN	392	446	PRO-RICH.
FT	DOMAIN	449	489	EGF-LIKE 6.
FT	DOMAIN	490	529	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	530	571	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	572	612	EGF-LIKE 9, CALCIUM-BINDING.
FT	DOMAIN	613	653	EGF-LIKE 10, CALCIUM-BINDING.
FT	REPEAT	654	722	TGFBP 2.
FT	DOMAIN	723	764	EGF-LIKE 11, CALCIUM-BINDING.
FT	DOMAIN	765	806	EGF-LIKE 12, CALCIUM-BINDING.
FT	DOMAIN	807	846	EGF-LIKE 13, CALCIUM-BINDING.
FT	DOMAIN	910	951	EGF-LIKE 14, CALCIUM-BINDING.
FT	REPEAT	952	1027	TGFBP 3.
FT	DOMAIN	1028	1069	EGF-LIKE 15, CALCIUM-BINDING.
FT	DOMAIN	1070	1112	EGF-LIKE 16, CALCIUM-BINDING.
FT	DOMAIN	1113	1154	EGF-LIKE 17, CALCIUM-BINDING.
FT	DOMAIN	1155	1196	EGF-LIKE 18, CALCIUM-BINDING.
FT	DOMAIN	1197	1237	EGF-LIKE 19, CALCIUM-BINDING.
FT	DOMAIN	1238	1279	EGF-LIKE 20, CALCIUM-BINDING.
FT	DOMAIN	1280	1321	EGF-LIKE 21, CALCIUM-BINDING.
FT	DOMAIN	1322	1362	EGF-LIKE 22, CALCIUM-BINDING.
FT	DOMAIN	1363	1403	EGF-LIKE 23, CALCIUM-BINDING.
FT	DOMAIN	1404	1445	EGF-LIKE 24, CALCIUM-BINDING.
FT	DOMAIN	1446	1486	EGF-LIKE 25, CALCIUM-BINDING.
FT	DOMAIN	1487	1527	EGF-LIKE 26, CALCIUM-BINDING.
FT	REPEAT	1528	1605	TGFBP 4.
FT	DOMAIN	1606	1647	EGF-LIKE 27, CALCIUM-BINDING.
FT	DOMAIN	1648	1688	EGF-LIKE 28, CALCIUM-BINDING.
FT	REPEAT	1689	1765	TGFBP 5.
FT	DOMAIN	1766	1807	EGF-LIKE 29, CALCIUM-BINDING.
FT	DOMAIN	1808	1848	EGF-LIKE 30, CALCIUM-BINDING.
FT	DOMAIN	1849	1890	EGF-LIKE 31, CALCIUM-BINDING.
FT	DOMAIN	1891	1929	EGF-LIKE 32, CALCIUM-BINDING.
FT	DOMAIN	1930	1972	EGF-LIKE 33, CALCIUM-BINDING.
FT	DOMAIN	1973	2012	EGF-LIKE 34, CALCIUM-BINDING.
FT	DOMAIN	2013	2054	EGF-LIKE 35, CALCIUM-BINDING.
FT	REPEAT	2055	2126	TGFBP 6.
FT	DOMAIN	2127	2165	EGF-LIKE 36, CALCIUM-BINDING.
FT	DOMAIN	2166	2205	EGF-LIKE 37, CALCIUM-BINDING.
FT	DOMAIN	2206	2246	EGF-LIKE 38, CALCIUM-BINDING.
FT	DOMAIN	2247	2290	EGF-LIKE 39, CALCIUM-BINDING.
FT	DOMAIN	2291	2332	EGF-LIKE 40, CALCIUM-BINDING.
FT	REPEAT	2333	2401	TGFBP 7.
FT	DOMAIN	2402	2443	EGF-LIKE 41, CALCIUM-BINDING.
FT	DOMAIN	2444	2484	EGF-LIKE 42, CALCIUM-BINDING.
FT	DOMAIN	2485	2523	EGF-LIKE 43, CALCIUM-BINDING.
FT	DOMAIN	2524	2566	EGF-LIKE 44, CALCIUM-BINDING.
FT	DOMAIN	2567	2606	EGF-LIKE 45, CALCIUM-BINDING.
FT	DOMAIN	2607	2647	EGF-LIKE 46, CALCIUM-BINDING.
FT	DOMAIN	2648	2687	EGF-LIKE 47, CALCIUM-BINDING.
FT	DISULFID	85	94	BY SIMILARITY.
FT	DISULFID	89	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	129	BY SIMILARITY.
FT	DISULFID	123	134	BY SIMILARITY.
FT	DISULFID	136	145	BY SIMILARITY.
FT	DISULFID	150	160	BY SIMILARITY.
FT	DISULFID	154	166	BY SIMILARITY.
FT	DISULFID	168	177	BY SIMILARITY.
FT	DISULFID	250	262	BY SIMILARITY.
FT	DISULFID	257	271	BY SIMILARITY.
FT	DISULFID	273	286	BY SIMILARITY.
FT	DISULFID	292	304	BY SIMILARITY.

```
FT   DISULFID   299    313       BY SIMILARITY.
FT   DISULFID   315    328       BY SIMILARITY.
FT   DISULFID   453    465       BY SIMILARITY.
FT   DISULFID   460    474       BY SIMILARITY.
FT   DISULFID   476    488       BY SIMILARITY.
FT   DISULFID   494    504       BY SIMILARITY.
FT   DISULFID   499    513       BY SIMILARITY.
FT   DISULFID   515    528       BY SIMILARITY.
FT   DISULFID   534    546       BY SIMILARITY.
FT   DISULFID   541    555       BY SIMILARITY.
FT   DISULFID   557    570       BY SIMILARITY.
FT   DISULFID   576    587       BY SIMILARITY.
FT   DISULFID   582    596       BY SIMILARITY.
FT   DISULFID   598    611       BY SIMILARITY.
FT   DISULFID   617    628       BY SIMILARITY.
FT   DISULFID   623    637       BY SIMILARITY.
FT   DISULFID   639    652       BY SIMILARITY.
FT   DISULFID   727    739       BY SIMILARITY.
FT   DISULFID   734    748       BY SIMILARITY.
FT   DISULFID   750    763       BY SIMILARITY.
FT   DISULFID   769    781       BY SIMILARITY.
FT   DISULFID   776    790       BY SIMILARITY.
FT   DISULFID   792    805       BY SIMILARITY.
FT   DISULFID   811    821       BY SIMILARITY.
FT   DISULFID   816    830       BY SIMILARITY.
FT   DISULFID   832    845       BY SIMILARITY.
FT   DISULFID   914    926       BY SIMILARITY.
FT   DISULFID   921    935       BY SIMILARITY.
FT   DISULFID   937    950       BY SIMILARITY.
FT   DISULFID  1032   1044       BY SIMILARITY.
FT   DISULFID  1039   1053       BY SIMILARITY.
FT   DISULFID  1055   1068       BY SIMILARITY.
FT   DISULFID  1074   1086       BY SIMILARITY.
FT   DISULFID  1081   1095       BY SIMILARITY.
FT   DISULFID  1097   1111       BY SIMILARITY.
FT   DISULFID  1117   1129       BY SIMILARITY.
FT   DISULFID  1124   1138       BY SIMILARITY.
FT   DISULFID  1140   1153       BY SIMILARITY.
FT   DISULFID  1159   1171       BY SIMILARITY.
FT   DISULFID  1201   1212       BY SIMILARITY.
FT   DISULFID  1208   1221       BY SIMILARITY.
FT   DISULFID  1223   1236       BY SIMILARITY.
FT   DISULFID  1242   1254       BY SIMILARITY.
FT   DISULFID  1249   1263       BY SIMILARITY.
FT   DISULFID  1265   1278       BY SIMILARITY.
FT   DISULFID  1284   1296       BY SIMILARITY.
FT   DISULFID  1291   1305       BY SIMILARITY.
FT   DISULFID  1307   1320       BY SIMILARITY.
FT   DISULFID  1326   1339       BY SIMILARITY.
FT   DISULFID  1333   1348       BY SIMILARITY.
FT   DISULFID  1350   1361       BY SIMILARITY.
FT   DISULFID  1367   1380       BY SIMILARITY.
FT   DISULFID  1374   1389       BY SIMILARITY.
FT   DISULFID  1391   1402       BY SIMILARITY.
FT   DISULFID  1408   1420       BY SIMILARITY.
FT   DISULFID  1415   1429       BY SIMILARITY.
FT   DISULFID  1450   1461       BY SIMILARITY.
FT   DISULFID  1456   1470       BY SIMILARITY.
FT   DISULFID  1472   1485       BY SIMILARITY.
FT   DISULFID  1491   1502       BY SIMILARITY.
FT   DISULFID  1497   1511       BY SIMILARITY.
FT   DISULFID  1513   1526       BY SIMILARITY.
FT   DISULFID  1610   1622       BY SIMILARITY.
FT   DISULFID  1617   1631       BY SIMILARITY.
FT   DISULFID  1633   1646       BY SIMILARITY.
FT   DISULFID  1652   1663       BY SIMILARITY.
FT   DISULFID  1658   1672       BY SIMILARITY.
FT   DISULFID  1674   1687       BY SIMILARITY.
FT   DISULFID  1770   1782       BY SIMILARITY.
FT   DISULFID  1777   1791       BY SIMILARITY.
FT   DISULFID  1793   1806       BY SIMILARITY.
FT   DISULFID  1812   1824       BY SIMILARITY.
FT   DISULFID  1818   1833       BY SIMILARITY.
FT   DISULFID  1835   1847       BY SIMILARITY.
FT   DISULFID  1853   1865       BY SIMILARITY.
FT   DISULFID  1860   1874       BY SIMILARITY.
FT   DISULFID  1876   1889       BY SIMILARITY.
FT   DISULFID  1895   1905       BY SIMILARITY.

  Query Match             6.7%;  Score 270;  DB 1;  Length 2871;
  Best Local Similarity  29.3%;  Pred. No. 4.7e-10;
  Matches   75;  Conservative   25;  Mismatches  102;  Indels   54;  Gaps   11;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  :| | | ||| : ||   |  |  | ||
Db       1289 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCDRHAVCTNTA 1343

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
              | | | |  |:              ||  |||:|||   : :|     | || ||: | | :
Db       1344 GSFNCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKE 1393

Qy        155 GYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVP 214
              ||           |  | :| ::|     |  :  :     | :  | |      ||
Db       1394 GY----------TGDGFTCADLD-----ECSENVKLCGNVQCLYAPGGYHCEYDMGFVP 1437

Qy        215 EDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFE--S 272
                 ||         | : |    | | |                |: |  |: |:|
Db       1438 SADRKSCVDSDECSLPNI-CVFGTCHNLP---------------GLFRCECEIGYELDR 1480

Qy        273 PGGKITSV--CTEKGT 286
               ||  | |  | | |
Db       1481 SGGNCTDVNECLEPPT 1496


RESULT 10
LTBS_HUMAN
ID   LTBS_HUMAN     STANDARD;      PRT;  1394 AA.
AC   P22064; Q8TD95;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Latent transforming growth factor beta binding protein, isoform 1S
DE   precursor (LTBP-1) (Transforming growth factor beta-1 binding protein
DE   1) (TGF-beta1-BP-1).
GN   LTBP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Fibroblast, and Platelet;
RX   MEDLINE=90275601; PubMed=2350783;
RA   Kanzaki T., Olofsson A., Moren A., Wernstedt C., Hellman U.,
RA   Miyazono K., Claesson-Welsh L., Heldin C.-H.;
RT   "TGF-beta 1 binding protein: a component of the large latent complex
RT   of TGF-beta 1 with multiple repeat sequences.";
RL   Cell 61:1051-1061(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kwak J.H., Shin K.Y., Kim S.I.;
RT   "Major alternative spliced-form of LTBP1 mRNA in human glomerular
RT   endothelial cell.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   INTERACTION WITH FIBRILLIN.
RX   PubMed=12429738;
RA   Isogai Z., Ono R.N., Ushiro S., Keene D.R., Chen Y., Mazzieri R.,
RA   Charbonneau N.L., Reinhardt D.P., Rifkin D.B., Sakai L.Y.;
RT   "Latent transforming growth factor beta-binding protein 1 interacts
RT   with fibrillin and is a microfibril-associated protein.";
RL   J. Biol. Chem. 278:2750-2757(2003).
CC   -!- SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC       composed of the TGF-beta1 molecule noncovalently associated with a
CC       disulfide-bonded complex of a dimer of the N-terminal propeptide
CC       of the TGF-beta1 precursor and a third component denoted TGF-
```

```
CC       beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC       Binds to fibrillin.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Short;
CC         IsoId=P22064-1; Sequence=Displayed;
CC       Name=Long;
CC         IsoId=Q14766-1; Sequence=External;
CC   -!- PTM: CONTAINS HYDROXYLATED ASPARAGINE RESIDUES.
CC   -!- PTM: The N-terminus is blocked.
CC   -!- SIMILARITY: Contains 16 EGF-like domains.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M34057; AAA61160.1; -.
DR   EMBL; AF489528; AAM03124.1; -.
DR   PIR; A35626; A35626.
DR   HSSP; P00750; 1TPG.
DR   GlycoSuiteDB; P22064; -.
DR   Genew; HGNC:6714; LTBP1.
DR   MIM; 150390; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   Pfam; PF00008; EGF; 15.
DR   Pfam; PF00683; TB; 4.
DR   SMART; SM00179; EGF_CA; 13.
DR   PROSITE; PS00010; ASX_HYDROXYL; 13.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 15.
KW   Growth factor binding; Repeat; EGF-like domain; Hydroxylation; Signal;
KW   Glycoprotein; Alternative splicing.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   1394       LATENT TRANSFORMING GROWTH FACTOR BETA
FT                                BINDING PROTEIN, ISOFORM 1S.
FT   DOMAIN      300    340       EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT      348    412       REPEAT A.
FT   DOMAIN      546    587       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      588    629       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      630    670       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      671    710       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      711    751       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      752    792       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      793    833       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      834    874       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      875    916       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      917    958       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      959   1001       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT     1017   1084       REPEAT B.
FT   DOMAIN     1097   1139       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT     1190   1262       REPEAT C.
FT   DOMAIN     1140   1180       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1294   1334       EGF-LIKE 15.
FT   DOMAIN     1335   1379       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   SITE        847    849       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    304    315       BY SIMILARITY.
FT   DISULFID    310    324       BY SIMILARITY.
FT   DISULFID    326    339       BY SIMILARITY.
FT   DISULFID    550    562       BY SIMILARITY.
FT   DISULFID    557    571       BY SIMILARITY.
FT   DISULFID    573    586       BY SIMILARITY.
FT   DISULFID    592    604       BY SIMILARITY.
FT   DISULFID    599    613       BY SIMILARITY.
FT   DISULFID    615    628       BY SIMILARITY.
FT   DISULFID    634    645       BY SIMILARITY.
FT   DISULFID    640    654       BY SIMILARITY.
FT   DISULFID    657    669       BY SIMILARITY.
FT   DISULFID    675    686       BY SIMILARITY.
FT   DISULFID    681    695       BY SIMILARITY.
FT   DISULFID    698    709       BY SIMILARITY.
FT   DISULFID    715    726       BY SIMILARITY.
FT   DISULFID    721    735       BY SIMILARITY.
FT   DISULFID    737    750       BY SIMILARITY.
FT   DISULFID    756    767       BY SIMILARITY.
FT   DISULFID    762    776       BY SIMILARITY.
FT   DISULFID    778    791       BY SIMILARITY.
FT   DISULFID    797    808       BY SIMILARITY.
FT   DISULFID    803    817       BY SIMILARITY.
FT   DISULFID    819    832       BY SIMILARITY.
FT   DISULFID    838    850       BY SIMILARITY.
FT   DISULFID    845    859       BY SIMILARITY.
FT   DISULFID    861    873       BY SIMILARITY.
FT   DISULFID    879    891       BY SIMILARITY.
FT   DISULFID    885    900       BY SIMILARITY.
FT   DISULFID    902    915       BY SIMILARITY.
FT   DISULFID    921    933       BY SIMILARITY.
FT   DISULFID    927    942       BY SIMILARITY.
FT   DISULFID    944    957       BY SIMILARITY.
FT   DISULFID    963    975       BY SIMILARITY.
FT   DISULFID    970    984       BY SIMILARITY.
FT   DISULFID    986   1000       BY SIMILARITY.
FT   DISULFID   1101   1114       BY SIMILARITY.
FT   DISULFID   1109   1123       BY SIMILARITY.
FT   DISULFID   1125   1138       BY SIMILARITY.
FT   DISULFID   1144   1155       BY SIMILARITY.
FT   DISULFID   1150   1164       BY SIMILARITY.
FT   DISULFID   1166   1179       BY SIMILARITY.
FT   DISULFID   1298   1309       BY SIMILARITY.
FT   DISULFID   1304   1318       BY SIMILARITY.
FT   DISULFID   1320   1333       BY SIMILARITY.
FT   DISULFID   1339   1354       BY SIMILARITY.
FT   DISULFID   1349   1363       BY SIMILARITY.
FT   DISULFID   1365   1378       BY SIMILARITY.
FT   MOD_RES     647    647       HYDROXYLATION.
FT   MOD_RES     810    810       HYDROXYLATION.
FT   CARBOHYD     21     21       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     52     52       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     98     98       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    294    294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    870    870       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    923    923       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1039   1039       N-LINKED (GLCNAC. . .).
FT                                /FTId=CAR_000184.
FT   CONFLICT    365    365       Y -> H (IN REF. 2).
FT   CONFLICT    504    504       A -> T (IN REF. 2).
FT   CONFLICT   1334   1334       F -> V (IN REF. 2).
SQ   SEQUENCE   1394 AA;  152791 MW;  DFFCA81A40B2C7D1 CRC64;
```

Query Match 6.5%; Score 260.5; DB 1; Length 1394;
Best Local Similarity 27.6%; Pred. No. 8.4e-10;
Matches 82; Conservative 32; Mismatches 102; Indels 81; Gaps 16;

```
Qy        34 DGLDVCATCH--EHATCQQREGKKICICNYGFVGNG-RTQCVDKNECQFGATLVCGNHTS 90
             : :| |    |   |  |:  ||   |:|: |:  :|    | | |||    : ||
Db       792 EDIDECQHRHLCAHGQCRNTEGSFQCVCDQGYRASGLGDHCEDINECLEDKS-VC-QRGD 849

Qy        91 CHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFEC 150
             | || | : | | :|::  ::|||        | ||:|||  |||   | |:|| ||| |
Db       850 CINTAGSYDCTCPDGFQ-LDDNKT--------CQDINECEHPGLCGPQGECLNTEGSFHC 900

Qy       151 YCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACREGF 210
             |  |          |  : |  :| :||      | | :    :  :      |  | :|
Db       901 VCQQG---------FSISADGRTCEDIDECVNNTVCDSH---GFCDNTAGSFRCLCYQG- 947

Qy       211 FSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGF 270
             |  |:|      | |        | ::                |   | |    || | |
Db       948 FQAPQD------GQG--------CVDV--------------NECELLSG----VCGEAF 974
```

```
Qy        271 -ESPGGKITSVCTEK--------GTWRESTLTCTEI------------LTKINDVSL 306
                |: :|    || ::        | | | | ::               :|| ||
Db        975 CENVEGSFLCVCADENQEYSPMTGQCRSRTSTDLDVDVDQPKEEKKECYYNLNDASL 1031


RESULT 11
LTBL_HUMAN
ID   LTBL_HUMAN     STANDARD;      PRT;  1595 AA.
AC   Q14766;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Latent transforming growth factor beta binding protein, isoform 1L
DE   precursor (LTBP-1) (Transforming growth factor beta-1 binding protein
DE   1) (TGF-beta1-BP-1).
GN   LTBP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE OF 1-346 FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=96125117; PubMed=8537398;
RA   Olofsson A., Ichijo H., Moren A., ten Dijke P., Miyazono K.,
RA   Heldin C.-H.;
RT   "Efficient association of an amino-terminally extended form of human
RT   latent transforming growth factor-beta binding protein with the
RT   extracellular matrix.";
RL   J. Biol. Chem. 270:31294-31297(1995).
RN   [2]
RP   SEQUENCE OF 347-1595 FROM N.A.
RC   TISSUE=Fibroblast, and Platelet;
RX   MEDLINE=90275601; PubMed=2350783;
RA   Kanzaki T., Olofsson A., Moren A., Wernstedt C., Hellman U.,
RA   Miyazono K., Claesson-Welsh L., Heldin C.-H.;
RT   "TGF-beta 1 binding protein: a component of the large latent complex
RT   of TGF-beta 1 with multiple repeat sequences.";
RL   Cell 61:1051-1061(1990).
RN   [3]
RP   INTERACTION WITH FIBRILLIN.
RX   PubMed=12429738;
RA   Isogai Z., Ono R.N., Ushiro S., Keene D R., Chen Y., Mazzieri R.,
RA   Charbonneau N.L., Reinhardt D.P., Rifkin D.B., Sakai L.Y.;
RT   "Latent transforming growth factor beta-binding protein 1 interacts
RT   with fibrillin and is a microfibril-associated protein.";
RL   J. Biol. Chem. 278:2750-2757(2003).
CC   -!- SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC       composed of the TGF-beta1 molecule noncovalently associated with a
CC       disulfide-bonded complex of a dimer of the N-terminal propeptide
CC       of the TGF-beta1 precursor and a third component denoted TGF-
CC       beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC       Binds to fibrillin.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=Q14766-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=P22064-1; Sequence=External;
CC   -!- TISSUE SPECIFICITY: The long isoform is found in fibroblasts.
CC   -!- PTM: CONTAINS HYDROXYLATED ASPARAGINE RESIDUES (BY SIMILARITY).
CC   -!- SIMILARITY: Contains 16 EGF-like domains.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L48925; AAA96327.1; -.
DR   EMBL; M34057; AAA61160.1; ALT_INIT.
DR   HSSP; P08709; 1BF9.
DR   MIM; 150390; -.
DR   GO; GO:0005578; C:extracellular matrix; NAS.
DR   GO; GO:0005024; F:transforming growth factor-beta receptor ac. . .; NAS.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   Pfam; PF00008; EGF; 15.
DR   Pfam; PF00683; TB; 4.
DR   SMART; SM00179; EGF_CA; 13.
DR   PROSITE; PS00010; ASX_HYDROXYL; 13.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 15.
KW   Growth factor binding; Repeat; EGF-like domain; Hydroxylation; Signal;
KW   Glycoprotein; Alternative splicing.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   1595       LATENT TRANSFORMING GROWTH FACTOR BETA
FT                                BINDING PROTEIN, ISOFORM 1L.
FT   DOMAIN      501    541       EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT      549    613       REPEAT A.
FT   DOMAIN      747    788       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      789    830       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      831    871       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      872    911       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      912    952       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      953    993       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      994   1034       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1035   1075       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1076   1117       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1118   1159       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1160   1202       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT     1218   1285       REPEAT B.
FT   DOMAIN     1298   1340       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT     1391   1463       REPEAT C.
FT   DOMAIN     1341   1381       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1495   1535       EGF-LIKE 15.
FT   DOMAIN     1536   1580       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   SITE       1048   1050       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    505    516       BY SIMILARITY.
FT   DISULFID    511    525       BY SIMILARITY.
FT   DISULFID    527    540       BY SIMILARITY.
FT   DISULFID    751    763       BY SIMILARITY.
FT   DISULFID    758    772       BY SIMILARITY.
FT   DISULFID    774    787       BY SIMILARITY.
FT   DISULFID    793    805       BY SIMILARITY.
FT   DISULFID    800    814       BY SIMILARITY.
FT   DISULFID    816    829       BY SIMILARITY.
FT   DISULFID    835    846       BY SIMILARITY.
FT   DISULFID    841    855       BY SIMILARITY.
FT   DISULFID    858    870       BY SIMILARITY.
FT   DISULFID    876    887       BY SIMILARITY.
FT   DISULFID    882    896       BY SIMILARITY.
FT   DISULFID    899    910       BY SIMILARITY.
FT   DISULFID    916    927       BY SIMILARITY.
FT   DISULFID    922    936       BY SIMILARITY.
FT   DISULFID    938    951       BY SIMILARITY.
FT   DISULFID    957    968       BY SIMILARITY.
FT   DISULFID    963    977       BY SIMILARITY.
FT   DISULFID    979    992       BY SIMILARITY.
FT   DISULFID    998   1009       BY SIMILARITY.
FT   DISULFID   1004   1018       BY SIMILARITY.
FT   DISULFID   1020   1033       BY SIMILARITY.
FT   DISULFID   1039   1051       BY SIMILARITY.
FT   DISULFID   1046   1060       BY SIMILARITY.
FT   DISULFID   1062   1074       BY SIMILARITY.
FT   DISULFID   1080   1092       BY SIMILARITY.
FT   DISULFID   1086   1101       BY SIMILARITY.
FT   DISULFID   1103   1116       BY SIMILARITY.
FT   DISULFID   1122   1134       BY SIMILARITY.
```

```
FT   DISULFID   1128   1143       BY SIMILARITY.
FT   DISULFID   1145   1158       BY SIMILARITY.
FT   DISULFID   1164   1176       BY SIMILARITY.
FT   DISULFID   1171   1185       BY SIMILARITY.
FT   DISULFID   1187   1201       BY SIMILARITY.
FT   DISULFID   1302   1315       BY SIMILARITY.
FT   DISULFID   1310   1324       BY SIMILARITY.
FT   DISULFID   1326   1339       BY SIMILARITY.
FT   DISULFID   1345   1356       BY SIMILARITY.
FT   DISULFID   1351   1365       BY SIMILARITY.
FT   DISULFID   1367   1380       BY SIMILARITY.
FT   DISULFID   1499   1510       BY SIMILARITY.
FT   DISULFID   1505   1519       BY SIMILARITY.
FT   DISULFID   1521   1534       BY SIMILARITY.
FT   DISULFID   1540   1555       BY SIMILARITY.
FT   DISULFID   1550   1564       BY SIMILARITY.
FT   DISULFID   1566   1579       BY SIMILARITY.
FT   MOD_RES     848    848       HYDROXYLATION (BY SIMILARITY).
FT   MOD_RES    1011   1011       HYDROXYLATION (BY SIMILARITY).
FT   CARBOHYD    495    495       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1071   1071       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1124   1124       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1240   1240       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1595 AA;  173229 MW;  6A091EBA8556D8E5 CRC64;

  Query Match             6.5%;  Score 260.5;  DB 1;  Length 1595;
  Best Local Similarity   27.6%;  Pred. No. 9.9e-10;
  Matches   82;  Conservative   32;  Mismatches  102;  Indels   81;  Gaps   16;

Qy        34 DGLDVCATCH--EHATCQQREGKKICICNYGFVGNG-RTQCVDKNECQFGATLVCGNHTS 90
             : :| |    |   |  |:  ||    |:|: |:  :|     | | |||     : ||
Db       993 EDIDECQHRHLCAHGQCRNTEGSFQCVCDQGYRASGLGDHCEDINECLEDKS-VC-QRGD 1050

Qy        91 CHNTPGGFYCICLEGYRATNNNKTFIPNDGTPCTDIDECEVSGLCRHGGRCVNTHGSFEC 150
             | || | : | | :|::  ::|||        | ||:|||  |||    | |:|| ||| |
Db       1051 CINTAGSYDCTCPDGFQ-LDDNKT--------CQDINECEHPGLCGPQGECLNTEGSFHC 1101

Qy       151 YCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACREGF 210
             |  |         |  : |  :| :||      | | :    :  :      |  | :|
Db       1102 VCQQG---------FSISADGRTCEDIDECVNNTVCDSH---GFCDNTAGSFRCLCYQG- 1148

Qy       211 FSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGF 270
             |  |:|      | |        | ::                |    | |    || | |
Db       1149 FQAPQD------GQG--------CVDV--------------NECELLSG----VCGEAF 1175

Qy       271 -ESPGGKITSVCTEK--------GTWRESTLPCTEI-----------LTKINDVSL 306
              |:  |    || ::        |  |  | |  ::                :|| ||
Db       1176 CENVEGSFLCVCADENQEYSPMTGQCRSRTSPDLDVDVDQPKEEKKECYYNLNDASL 1232


RESULT 12
EMR1_MOUSE
ID   EMR1_MOUSE     STANDARD;      PRT;   931 AA.
AC   Q61549;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Cell surface glycoprotein EMR1 precursor (EMR1 hormone receptor)
DE   (Cell surface glycoprotein F4/80).
GN   EMR1 OR GPF480.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Peritoneal cavity;
RX   MEDLINE=96132946; PubMed=8550607;
RA   McKnight A.J., Macfarlane A.J., Dri P., Turley L., Willis A.C.,
RA   Gordon S.;
RT   "Molecular cloning of F4/80, a murine macrophage-restricted cell
RT   surface glycoprotein with homology to the G-protein-linked
RT   transmembrane 7 hormone receptor family.";
RL   J. Biol. Chem. 271:486-489(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97312684; PubMed=9169125;
RA   Lin H.H., Stubbs L.J., Mucenski M.L.;
RT   "Identification and characterization of a seven transmembrane hormone
RT   receptor using differential display.";
RL   Genomics 41:301-308(1997).
CC   -!- FUNCTION: PROBABLY INVOLVED IN CELL ADHESION WITHIN TISSUES
CC       AND RECEPTOR SIGNALLING.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: IN MACROPHAGES; BUT ABSENT FROM THOSE WHICH
CC       ARE LOCALIZED WITHIN T-CELL AREAS OF LYMPH NODES AND SPLEEN.
CC       LOW LEVEL OF EXPRESSION ON BLOOD MONOCYTES.
CC   -!- SIMILARITY: Contains 7 EGF-like domains.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: Contains 1 GPS domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X93328; CAA63720.1; -.
DR   EMBL; U66888; AAC53184.1; -.
DR   HSSP; P35555; 1EMN.
DR   MGD; MGI:106912; Emr1.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF01825; GPS; 1.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00179; EGF_CA; 6.
DR   SMART; SM00303; GPS; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 6.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 5.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
KW   Cell adhesion; Receptor; G-protein coupled receptor; Repeat; Signal;
KW   Transmembrane; EGF-like domain; Glycoprotein.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    931       CELL SURFACE GLYCOPROTEIN EMR1.
FT   DOMAIN       28    644       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    645    672       POTENTIAL.
FT   DOMAIN      673    679       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    680    701       POTENTIAL.
FT   DOMAIN      702    711       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    712    735       POTENTIAL.
FT   DOMAIN      736    754       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    755    776       POTENTIAL.
FT   DOMAIN      777    792       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    793    821       POTENTIAL.
FT   DOMAIN      822    839       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    840    859       POTENTIAL.
FT   DOMAIN      860    874       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    875    897       POTENTIAL.
FT   DOMAIN      898    931       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       32     80       EGF-LIKE 1.
FT   DOMAIN       81    132       EGF-LIKE 2.
FT   DOMAIN      133    172       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      173    221       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      222    271       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      272    318       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      319    367       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
```

```
FT   DOMAIN      592    641       GPS.
FT   SITE        506    508       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID     36     48       BY SIMILARITY.
FT   DISULFID     42     57       BY SIMILARITY.
FT   DISULFID     59     79       BY SIMILARITY.
FT   DISULFID     85     98       BY SIMILARITY.
FT   DISULFID     92    107       BY SIMILARITY.
FT   DISULFID    109    131       BY SIMILARITY.
FT   DISULFID    137    149       BY SIMILARITY.
FT   DISULFID    143    158       BY SIMILARITY.
FT   DISULFID    160    171       BY SIMILARITY.
FT   DISULFID    177    189       BY SIMILARITY.
FT   DISULFID    183    198       BY SIMILARITY.
FT   DISULFID    200    220       BY SIMILARITY.
FT   DISULFID    226    239       BY SIMILARITY.
FT   DISULFID    233    248       BY SIMILARITY.
FT   DISULFID    250    270       BY SIMILARITY.
FT   DISULFID    276    286       BY SIMILARITY.
FT   DISULFID    280    295       BY SIMILARITY.
FT   DISULFID    297    317       BY SIMILARITY.
FT   DISULFID    323    336       BY SIMILARITY.
FT   DISULFID    330    345       BY SIMILARITY.
FT   DISULFID    347    366       BY SIMILARITY.
FT   CARBOHYD    148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    167    167       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    229    229       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    283    283       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    405    405       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    417    417       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    474    474       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    498    498       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    706    706       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   931 AA;  102129 MW;  52963A667E8B76B5 CRC64;

  Query Match             6.5%;  Score 260;  DB 1;  Length 931;
  Best Local Similarity   23.9%;  Pred. No. 5.4e-10;
  Matches  123;  Conservative   54;  Mismatches  192;  Indels  146;  Gaps   27;

Qy        20 LLGLA-----RGAAGAPGPD-----------GLDVCAT---CHEHATCQQREGKKICIC 59
             :|| |      || : : | |             :| | |   | ::: |    |   | ||
Db       101 ILGRAKCSCLRGFSSSTGKDWILGSLDNFLCADVDECLTIGICPKYSNCSNSVGSYSCTC 160

Qy        60 NYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNN-------- 111
             ||| || : | |::||      ||  | :|||| | :|| |  |  ::
Db       161 QPGFVLNG-SICEDEDECV--TRDVCPEHATCHNTLGSYYCTCNSGLESSGGGPMFQGLD 217

Qy       112 ----------------NKTFI-------------------------PNDGTFCTDIDEC 129
                            |||                         | ||  |||||||
Db       218 ESCEDVDECSRNSTLCGPTFICINTLGSYSCSCPAGFSLPTFQILGHPADGN-CTDIDEC 276

Qy       130 EVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFH----PTTDATSCTE--IDCG--- 180
             :   |    | || ||: | |  |:   ||  |         |   ||:  : ||
Db       277 D--DTCPLNSSCTNTIGSYFCTCHPGFASSNGQLNFKDLEVTCEDIDECTQDPLQCGLNS 334

Qy       181 TPPEVPDGYIIG----NYTSSLGSQ--VRYACREGFFSVPEDTVSSCTGLGTWESPKL-H 233
                 ||  || |          |||    : |:   |   || :         :| ::
Db       335 VCTNVPGSYICGCLPDFQMDPEGSQGYGNFNCKRILFKCKEDLI--------LQSEQIQQ 386

Qy       234 CQEINCGNPPEMRHAILVGNHSSRLGGVARVCQE-----GFESPGGKITSVCTEKGTW- 287
             || :   :          ||    : |       |:       :|    :: |  :  ||
Db       387 CQAVQGRDLGYASFCTLVNATFTILDN---TCENKSAPVSLQSAATSVSLVLEQATTWF 442

Qy       288 ---RESTLTCTEILTKINDVSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVRE 344
                :| | |   || :  : ::      :      |: |  :    : |: : ::  |
Db       443 ELSKEETSTLGTILLETVESTMLAALLIPSGNASQMIQTE---YLDIESKVINE-ECKEN 498

Qy       345 ETVNLTTDSRTPEV-CLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDL----LEDDG 399
             |::||       | |  :        ::||  |     | :  |   |  |      ||
Db       499 ESINLAARGDKMNVGCFII------KESVSTGAPG---VAFVSFAHMESVLNERFFEDGQ 549

Qy       400 SFNISIFNETCLKLNRRSRKVGSEHMYQFTVLGQR 434
             ||           ||   || ||       || |::
Db       550 SFR---------KLRMNSRVVGG------TVTGEK 569


RESULT 13
LTBS_MOUSE
ID   LTBS_MOUSE     STANDARD;      PRT;  1389 AA.
AC   Q8CGI8; Q8BNW7; Q8C7F5; Q8CIR0;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Latent transforming growth factor beta binding protein, isoform 1S
DE   precursor (LTBP-1) (Transforming growth factor beta-1 binding protein
DE   1) (TGF-beta1-BP-1).
GN   LTBP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12711388;
RA   Noguera I., Obata H., Gualandris A., Cowin P., Rifkin D.B.;
RT   "Molecular cloning of the mouse Ltbp-1 gene reveals tissue specific
RT   expression of alternatively spliced forms.";
RL   Gene 308:31-41(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ;
RX   PubMed=12711389;
RA   Weiskirchen R., Moser M., Guenther K., Weiskirchen S., Gressner A.M.;
RT   "The murine latent transforming growth factor-beta binding protein
RT   (Ltbp-1) is alternatively spliced, and maps to a region syntenic to
RT   human chromosome 2p21-22.";
RL   Gene 308:43-52(2003).
RN   [3]
RP   SEQUENCE OF 788-1389 FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Aorta, Liver, and Vein;
RX   PubMed=12466851;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
CC   -!- SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC       composed of the TGF-beta1 molecule noncovalently associated with a
CC       disulfide-bonded complex of a dimer of the N-terminal propeptide
CC       of the TGF-beta1 precursor and a third component denoted TGF-
CC       beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC       Binds to fibrillin (By similarity).
```

```
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1S;
CC         IsoId=Q8CG18-1; Sequence=Displayed;
CC       Name=1L;
CC         IsoId=Q8CG19-1; Sequence=External;
CC   -!- PTM: Contains hydroxylated asparagine residues (By similarity).
CC   -!- PTM: The N-terminus is blocked (By similarity).
CC   -!- SIMILARITY: Contains 16 EGF-like domains.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF346465; AAN77251.1; -.
DR   EMBL; AF346438; AAN77251.1; JOINED.
DR   EMBL; AF346439; AAN77251.1; JOINED.
DR   EMBL; AF346440; AAN77251.1; JOINED.
DR   EMBL; AF346441; AAN77251.1; JOINED.
DR   EMBL; AF346442; AAN77251.1; JOINED.
DR   EMBL; AF346443; AAN77251.1; JOINED.
DR   EMBL; AF346444; AAN77251.1; JOINED.
DR   EMBL; AF346445; AAN77251.1; JOINED.
DR   EMBL; AF346446; AAN77251.1; JOINED.
DR   EMBL; AF346447; AAN77251.1; JOINED.
DR   EMBL; AF346448; AAN77251.1; JOINED.
DR   EMBL; AF346449; AAN77251.1; JOINED.
DR   EMBL; AF346450; AAN77251.1; JOINED.
DR   EMBL; AF346451; AAN77251.1; JOINED.
DR   EMBL; AF346452; AAN77251.1; JOINED.
DR   EMBL; AF346453; AAN77251.1; JOINED.
DR   EMBL; AF346454; AAN77251.1; JOINED.
DR   EMBL; AF346455; AAN77251.1; JOINED.
DR   EMBL; AF346456; AAN77251.1; JOINED.
DR   EMBL; AF346457; AAN77251.1; JOINED.
DR   EMBL; AF346458; AAN77251.1; JOINED.
DR   EMBL; AF346459; AAN77251.1; JOINED.
DR   EMBL; AF346460; AAN77251.1; JOINED.
DR   EMBL; AF346461; AAN77251.1; JOINED.
DR   EMBL; AF346462; AAN77251.1; JOINED.
DR   EMBL; AF346463; AAN77251.1; JOINED.
DR   EMBL; AF346464; AAN77251.1; JOINED.
DR   EMBL; AY143161; AAN38831.1; ALT_SEQ.
DR   EMBL; AK050380; BAC34222.1; -.
DR   EMBL; AK080024; BAC37808.1; -.
DR   MGD; MGI:109151; Ltbp1.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 15.
DR   Pfam; PF00683; TB; 4.
DR   SMART; SM00181; EGF; 17.
DR   SMART; SM00179; EGF_CA; 16.
DR   PROSITE; PS00010; ASX_HYDROXYL; 13.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 15.
KW   Growth factor binding; Repeat; EGF-like domain;
KW   Hydroxylation; Signal; Glycoprotein;
KW   Alternative splicing.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1389       LATENT TRANSFORMING GROWTH FACTOR BETA
FT                                BINDING PROTEIN, ISOFORM 1S.
FT   DOMAIN       295    335       EGF-LIKE 1, CALCIUM-BINDING
FT                                (POTENTIAL)
FT   REPEAT       343    408       REPEAT A.
FT   DOMAIN       542    583       EGF-LIKE 2, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       584    625       EGF-LIKE 3, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       626    666       EGF-LIKE 4, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       667    706       EGF-LIKE 5, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       707    747       EGF-LIKE 6, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       748    788       EGF-LIKE 7, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       789    829       EGF-LIKE 8, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       830    870       EGF-LIKE 9, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       871    912       EGF-LIKE 10, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       913    954       EGF-LIKE 11, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN       955    997       EGF-LIKE 12, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   REPEAT      1012   1079       REPEAT B.
FT   DOMAIN      1092   1134       EGF-LIKE 13, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DOMAIN      1135   1175       EGF-LIKE 14, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   REPEAT      1185   1257       REPEAT C.
FT   DOMAIN      1289   1330       EGF-LIKE 15.
FT   DOMAIN      1330   1374       EGF-LIKE 16, CALCIUM-BINDING
FT                                (POTENTIAL).
FT   DISULFID     299    310       POTENTIAL.
FT   DISULFID     305    319       POTENTIAL.
FT   DISULFID     321    334       POTENTIAL.
FT   DISULFID     546    558       POTENTIAL.
FT   DISULFID     553    567       POTENTIAL.
FT   DISULFID     569    582       POTENTIAL.
FT   DISULFID     588    600       POTENTIAL.
FT   DISULFID     595    609       POTENTIAL.
FT   DISULFID     611    624       POTENTIAL.
FT   DISULFID     630    641       POTENTIAL.
FT   DISULFID     636    650       POTENTIAL.
FT   DISULFID     653    665       POTENTIAL.
FT   DISULFID     671    682       POTENTIAL.
FT   DISULFID     677    691       POTENTIAL.
FT   DISULFID     694    705       POTENTIAL.
FT   DISULFID     711    722       POTENTIAL.
FT   DISULFID     717    731       POTENTIAL.
FT   DISULFID     733    746       POTENTIAL.
FT   DISULFID     752    763       POTENTIAL.
FT   DISULFID     758    772       POTENTIAL.
FT   DISULFID     774    787       POTENTIAL.
FT   DISULFID     793    804       POTENTIAL.
FT   DISULFID     799    813       POTENTIAL.
FT   DISULFID     815    828       POTENTIAL.
FT   DISULFID     834    846       POTENTIAL.
FT   DISULFID     841    855       POTENTIAL.
FT   DISULFID     857    869       POTENTIAL.
FT   DISULFID     875    887       POTENTIAL.

  Query Match          6.4%;  Score 256;  DB 1;  Length 1389;
  Best Local Similarity  26.4%;  Pred. No. 1.7e-09;
  Matches   79;  Conservative   31;  Mismatches  103;  Indels   86;  Gaps   14;

Qy         34 DGLDVCATCH--EHATCQQREGKKICICNYGF----VGNGRTQCVDKNECQFGATLVCGN 87
              : :| |    |   |  |:  ||   |:|| |:    :|:    | | |||   ::: |
Db        788 EDIDECEHHHLCSHGQCRNTEGSFQCVCNQGYRASVLGD---HCEDINECLEDSSVCQGG 844

Qy         88 HTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGS 147
                | || | : | | :|::          ||    | ||:||   |||   | |:|| ||
Db        845 --DCINTAGSYDCTCPDGFQL---------NDNKGCQDINECAQPGLCGSHGECLNTQGS 893

Qy        148 FECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACR 207
```

```
Db        894 FHCVCEQG---------FSISADGRTCEDIDECVNNTVCDSH---GFCDNTAGSFRCLCY 941

Qy        208 EGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQ 267

Db        942 QG-FQAPQD------GQG--------CVDV-------------NECELLSG----VCG 967

Qy        268 EGF-ESPGGKITSVCTEK-------------------GTWRESTLTCTEILTKINDVSL 306

Db        968 EAFCENVEGSFLCVCADENQEYSPMTGQCRSRVTEDSGVDRQPREEKKECYYNLNDASL 1026
```

RESULT 14
LTBL_MOUSE
ID LTBL_MOUSE STANDARD; PRT; 1713 AA.
AC Q8CG19; O88349; Q8BNW7; Q8C7F5; Q8CIR0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Latent transforming growth factor beta binding protein, isoform 1L
DE precursor (LTBP-1) (Transforming growth factor beta-1 binding protein
DE 1) (TGF-beta1-BP-1).
GN LTBP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12711388;
RA Noguera I., Obata H., Gualandris A., Cowin P., Rifkin D.B.;
RT "Molecular cloning of the mouse Ltbp-1 gene reveals tissue specific
RT expression of alternatively spliced forms.";
RL Gene 308:31-41(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ;
RX PubMed=12711389;
RA Weiskirchen R., Moser M., Guenther K., Weiskirchen S., Gressner A.M.;
RT "The murine latent transforming growth factor-beta binding protein
RT (Ltbp-1) is alternatively spliced, and maps to a region syntenic to
RT human chromosome 2p21-22.";
RL Gene 308:43-52(2003).
RN [3]
RP SEQUENCE OF 1112-1713 FROM N.A.
RC STRAIN=C57BL/6J;
RC TISSUE=Aorta, Liver, and Vein;
RX PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
CC -!- SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC composed of the TGF-beta1 molecule noncovalently associated with a
CC disulfide-bonded complex of a dimer of the N-terminal propeptide
CC of the TGF-beta1 precursor and a third component denoted TGF-
CC beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC Binds to fibrillin (By similarity).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1L;
CC IsoId=Q8CG19-1; Sequence=Displayed;
CC Name=1S;
CC IsoId=Q8CG18-1; Sequence=External;
CC -!- PTM: Contains hydroxylated asparagine residues (By similarity).
CC -!- SIMILARITY: Contains 16 EGF-like domains.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; AF346465; AAN77250.1; -.
DR EMBL; AF346434; AAN77250.1; JOINED.
DR EMBL; AF346435; AAN77250.1; JOINED.
DR EMBL; AF346436; AAN77250.1; JOINED.
DR EMBL; AF346437; AAN77250.1; JOINED.
DR EMBL; AF346438; AAN77250.1; JOINED.
DR EMBL; AF346439; AAN77250.1; JOINED.
DR EMBL; AF346440; AAN77250.1; JOINED.
DR EMBL; AF346441; AAN77250.1; JOINED.
DR EMBL; AF346442; AAN77250.1; JOINED.
DR EMBL; AF346443; AAN77250.1; JOINED.
DR EMBL; AF346444; AAN77250.1; JOINED.
DR EMBL; AF346445; AAN77250.1; JOINED.
DR EMBL; AF346446; AAN77250.1; JOINED.
DR EMBL; AF346447; AAN77250.1; JOINED.
DR EMBL; AF346448; AAN77250.1; JOINED.
DR EMBL; AF346449; AAN77250.1; JOINED.
DR EMBL; AF346450; AAN77250.1; JOINED.
DR EMBL; AF346451; AAN77250.1; JOINED.
DR EMBL; AF346452; AAN77250.1; JOINED.
DR EMBL; AF346453; AAN77250.1; JOINED.
DR EMBL; AF346454; AAN77250.1; JOINED.
DR EMBL; AF346455; AAN77250.1; JOINED.
DR EMBL; AF346456; AAN77250.1; JOINED.
DR EMBL; AF346457; AAN77250.1; JOINED.
DR EMBL; AF346458; AAN77250.1; JOINED.
DR EMBL; AF346459; AAN77250.1; JOINED.
DR EMBL; AF346460; AAN77250.1; JOINED.
DR EMBL; AF346461; AAN77250.1; JOINED.
DR EMBL; AF346462; AAN77250.1; JOINED.
DR EMBL; AF346463; AAN77250.1; JOINED.
DR EMBL; AF346464; AAN77250.1; JOINED.
DR EMBL; AF022889; AAC33307.1; -.
DR EMBL; AY143161; AAN38831.1; ALT_SEQ.
DR EMBL; AK050380; BAC34222.1; -.
DR EMBL; AK080024; BAC37808.1; -.
DR HSSP; P08709; 1BF9.
DR MGD; MGI:109151; Ltbp1.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR002212; Fibril-assoc.
DR InterPro; IPR006210; IEGF.
DR Pfam; PF00008; EGF; 16.
DR Pfam; PF00683; TB; 4.

```
DR   SMART; SM00181; EGF; 18.
DR   SMART; SM00179; EGF_CA; 16.
DR   PROSITE; PS00010; ASX_HYDROXYL; 13.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 11.
DR   PROSITE; PS01187; EGF_CA; 15.
KW   Growth factor binding; Repeat; EGF-like domain;
KW   Hydroxylation; Signal; Glycoprotein;
KW   Alternative splicing.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24   1713       LATENT TRANSFORMING GROWTH FACTOR BETA
FT                                BINDING PROTEIN, ISOFORM 1L.
FT   DOMAIN      619    659       EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT      667    732       REPEAT A.
FT   DOMAIN      866    907       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      908    949       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      950    990       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      991   1030       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1031   1071       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1072   1112       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1113   1153       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1154   1194       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1195   1236       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1237   1278       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1279   1321       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT     1336   1403       REPEAT B.
FT   DOMAIN     1416   1458       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1459   1499       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT     1509   1581       REPEAT C.
FT   DOMAIN     1613   1653       EGF-LIKE 15.
FT   DOMAIN     1654   1698       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DISULFID    623    634       POTENTIAL.
FT   DISULFID    629    643       POTENTIAL.
FT   DISULFID    645    658       POTENTIAL.
FT   DISULFID    870    882       POTENTIAL.
FT   DISULFID    877    891       POTENTIAL.
FT   DISULFID    893    906       POTENTIAL.
FT   DISULFID    912    924       POTENTIAL.
FT   DISULFID    919    933       POTENTIAL.
FT   DISULFID    935    948       POTENTIAL.
FT   DISULFID    954    965       POTENTIAL.
FT   DISULFID    960    974       POTENTIAL.
FT   DISULFID    977    989       POTENTIAL.
FT   DISULFID    995   1006       POTENTIAL.
FT   DISULFID   1001   1015       POTENTIAL.
FT   DISULFID   1018   1029       POTENTIAL.
FT   DISULFID   1035   1046       POTENTIAL.
FT   DISULFID   1041   1055       POTENTIAL.
FT   DISULFID   1057   1070       POTENTIAL.
FT   DISULFID   1076   1087       POTENTIAL.
FT   DISULFID   1082   1096       POTENTIAL.
FT   DISULFID   1098   1111       POTENTIAL.
FT   DISULFID   1117   1128       POTENTIAL.
FT   DISULFID   1123   1137       POTENTIAL.
FT   DISULFID   1139   1152       POTENTIAL.
FT   DISULFID   1158   1170       POTENTIAL.
FT   DISULFID   1165   1179       POTENTIAL.
FT   DISULFID   1181   1193       POTENTIAL.
FT   DISULFID   1199   1211       POTENTIAL.
FT   DISULFID   1205   1220       POTENTIAL.
FT   DISULFID   1222   1235       POTENTIAL.
FT   DISULFID   1241   1253       POTENTIAL.
FT   DISULFID   1247   1262       POTENTIAL.
FT   DISULFID   1264   1277       POTENTIAL.
FT   DISULFID   1283   1295       POTENTIAL.
FT   DISULFID   1290   1304       POTENTIAL.
FT   DISULFID   1306   1320       POTENTIAL.
FT   DISULFID   1420   1433       POTENTIAL.

  Query Match             6.4%;  Score 256;  DB 1;  Length 1713;
  Best Local Similarity  26.4%;  Pred. No. 2.2e-09;
  Matches   79;  Conservative   31;  Mismatches  103;  Indels   86;  Gaps   14;
```

```
Qy         34 DGLDVCATCH--EHATCQQREGKKICICNYGF----VGNGRTQCVDKNECQFGATLVCGN 87
              : :| |    |    |  |:  ||   |:|| |:   :|:    | | |||   ::: |
Db       1112 EDIDECEHHHLCSHGQCRNTEGSFQCVCNQGYRASVLGD---HCEDINECLEDSSVCQGG 1168

Qy         88 HTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGS 147
              | || | : | | :|::          ||    | ||:||   |||   | |:|| ||
Db       1169 --DCINTAGSYDCTCPDGFQL---------NDNKGCQDINECAQPGLCGSHGECLNTQGS 1217

Qy        148 FECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACR 207
              | | |  |          | : |  :| :||      | | :    :  :      | |
Db       1218 FHCVCEQG---------FSISADGRTCEDIDECVNNTVCDSH---GFCDNTAGSFRCLCY 1265

Qy        208 EGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQ 267
              :| |  |:|     | |       | ::             |   | |    ||
Db       1266 QG-FQAPQD------GQG--------CVDV--------------NECELLSG----VCG 1291

Qy        268 EGF-ESPGGKITSVCTEK-----------------GTWRESTLTCTEILTKINDVSL 306
              | | |:  |    || ::                |  |:      |    :|| ||
Db       1292 EAFCENVEGSFLCVCADENQEYSPMTGQCRSRVTEDSGVDRQPREEKKECYYNLNDASL 1350
```

```
RESULT 15
LTB1_RAT
ID   LTB1_RAT       STANDARD;      PRT;  1712 AA.
AC   Q00918;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Latent transforming growth factor beta binding protein 1 precursor
DE   (LTBP-1) (Transforming growth factor beta-1 binding protein 1) (TGF-
DE   beta1-BP-1) (Transforming growth factor beta-1 masking protein, large
DE   subunit).
GN   LTBP1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=91062373; PubMed=2247454;
RA   Tsuji T., Okada F., Yamaguchi K., Nakamura T.;
RT   "Molecular cloning of the large subunit of transforming growth factor
RT   type beta masking protein and expression of the mRNA in various rat
RT   tissues.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:8835-8839(1990).
CC   -!- SUBUNIT: THE LARGE LATENT COMPLEX OF TGF-BETA1 FROM PLATELETS IS
CC       COMPOSED OF THE TGF-BETA1 MOLECULE NONCOVALENTLY ASSOCIATED WITH
CC       A MASKING PROTEIN CONSISTING OF A DISULFIDE-BONDED COMPLEX OF A
CC       DIMER OF THE N-TERMINAL PROPEPTIDE OF THE TGF-BETA1 PRECURSOR AND
CC       A THIRD COMPONENT DENOTED TGF-BETA1-BP (OR MP LARGE SUBUNIT).
CC       TGF-BETA1-BP DOES NOT BIND DIRECTLY TO ACTIVE TGF-BETA1.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: Contains 18 EGF-like domains.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M55431; AAA42235.1; -.
DR   PIR; A38261; A38261.
DR   HSSP; P16109; 1FSB.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002212; Fibril-assoc.
DR   Pfam; PF00008; EGF; 16.
DR   Pfam; PF00683; TB; 4.
DR   SMART; SM00179; EGF_CA; 13.
DR   PROSITE; PS00010; ASX_HYDROXYL; 13.
```

```
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 10.
DR   PROSITE; PS01187; EGF_CA; 15.
KW   Growth factor binding; Repeat; EGF-like domain; Hydroxylation; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21    736       POTENTIAL.
FT   CHAIN       737   1577       LATENT TRANSFORMING GROWTH FACTOR BETA
FT                                BINDING PROTEIN 1.
FT   SITE        734    736       CLEAVAGE (POTENTIAL).
FT   SITE       1575   1577       CLEAVAGE (POTENTIAL).
FT   PROPEP     1578   1712       POTENTIAL.
FT   DOMAIN      181    213       EGF-LIKE 1.
FT   DOMAIN      391    423       EGF-LIKE 2.
FT   REPEAT      551    604       INTERNAL REPEAT 1.
FT   DOMAIN      618    658       EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT      671    721       INTERNAL REPEAT 2.
FT   DOMAIN      865    906       EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      907    948       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      949    989       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      990   1029       EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1030   1070       EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1071   1111       EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1112   1152       EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1153   1193       EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1194   1235       EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1236   1277       EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1278   1320       EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT     1340   1392       INTERNAL REPEAT 3.
FT   DOMAIN     1415   1457       EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     1458   1498       EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   REPEAT     1517   1568       INTERNAL REPEAT 4.
FT   DOMAIN     1612   1652       EGF-LIKE 17.
FT   DOMAIN     1653   1697       EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT   DISULFID    185    195       BY SIMILARITY.
FT   DISULFID    189    201       BY SIMILARITY.
FT   DISULFID    203    212       BY SIMILARITY.
FT   DISULFID    395    405       BY SIMILARITY.
FT   DISULFID    399    411       BY SIMILARITY.
FT   DISULFID    413    422       BY SIMILARITY.
FT   DISULFID    622    633       BY SIMILARITY.
FT   DISULFID    628    642       BY SIMILARITY.
FT   DISULFID    644    657       BY SIMILARITY.
FT   DISULFID    869    881       BY SIMILARITY.
FT   DISULFID    876    890       BY SIMILARITY.
FT   DISULFID    892    905       BY SIMILARITY.
FT   DISULFID    911    923       BY SIMILARITY.
FT   DISULFID    918    932       BY SIMILARITY.
FT   DISULFID    934    947       BY SIMILARITY.
FT   DISULFID    953    964       BY SIMILARITY.
FT   DISULFID    959    973       BY SIMILARITY.
FT   DISULFID    976    988       BY SIMILARITY.
FT   DISULFID    994   1005       BY SIMILARITY.
FT   DISULFID   1000   1014       BY SIMILARITY.
FT   DISULFID   1017   1028       BY SIMILARITY.
FT   DISULFID   1034   1045       BY SIMILARITY.
FT   DISULFID   1040   1054       BY SIMILARITY.
FT   DISULFID   1056   1069       BY SIMILARITY.
FT   DISULFID   1075   1086       BY SIMILARITY.
FT   DISULFID   1081   1095       BY SIMILARITY.
FT   DISULFID   1097   1110       BY SIMILARITY.
FT   DISULFID   1116   1127       BY SIMILARITY.
FT   DISULFID   1122   1136       BY SIMILARITY.
FT   DISULFID   1138   1151       BY SIMILARITY.
FT   DISULFID   1157   1169       BY SIMILARITY.
FT   DISULFID   1164   1178       BY SIMILARITY.
FT   DISULFID   1180   1192       BY SIMILARITY.
FT   DISULFID   1198   1210       BY SIMILARITY.
FT   DISULFID   1204   1219       BY SIMILARITY.
FT   DISULFID   1221   1234       BY SIMILARITY.
FT   DISULFID   1240   1252       BY SIMILARITY.
FT   DISULFID   1246   1261       BY SIMILARITY.
FT   DISULFID   1263   1276       BY SIMILARITY.
FT   DISULFID   1282   1294       BY SIMILARITY.
FT   DISULFID   1289   1303       BY SIMILARITY.
FT   DISULFID   1305   1319       BY SIMILARITY.
FT   DISULFID   1419   1432       BY SIMILARITY.
FT   DISULFID   1427   1441       BY SIMILARITY.
FT   DISULFID   1443   1456       BY SIMILARITY.
FT   DISULFID   1462   1473       BY SIMILARITY.
FT   DISULFID   1468   1482       BY SIMILARITY.
FT   DISULFID   1484   1497       BY SIMILARITY.
FT   DISULFID   1616   1627       BY SIMILARITY.
FT   DISULFID   1622   1636       BY SIMILARITY.
FT   DISULFID   1638   1651       BY SIMILARITY.
FT   DISULFID   1657   1672       BY SIMILARITY.
FT   DISULFID   1667   1681       BY SIMILARITY.
FT   DISULFID   1683   1696       BY SIMILARITY.
FT   CARBOHYD    339    339       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    370    370       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    416    416       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1042   1042       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1242   1242       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1357   1357       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1712 AA;  186598 MW;  650BCEAA691FD134 CRC64;

  Query Match             6.2%;  Score 251;  DB 1;  Length 1712;
  Best Local Similarity   26.4%;  Pred. No. 4.8e-09;
  Matches   79;  Conservative   30;  Mismatches  104;  Indels   86;  Gaps   14;

Qy         34 DGLDVCATCH--EHATCQQREGKKICICNYGF----VGNGRTQCVDKNECQFGATLVCGN 87
              : :| |    |    |  |: ||    |:|| |:    :|:    | | |||   ::: |
Db       1111 EDIDECEHRHLCSHGQCRNTEGSFQCLCNQGYRASVLGD---HCEDINECLEDSSVCQGG 1167

Qy         88 HTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGS 147
                | || | : | | :| :          ||   | ||:||   |||   | |:|| ||
Db       1168 --DCINTAGSYDCTCPDGLQL---------NDNKGCQDINECAQPGLCAPHGECLNTQGS 1216

Qy        148 FECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACR 207
              | | |  |           |  : |  :| :||       | | :    :  :      |  |
Db       1217 FHCVCEQG---------FSISADGRTCEDIDECVNNTVCDSH---GFCDNTAGSFRCLCY 1264

Qy        208 EGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQ 267
              :| |  |:|      | |        | ::               |    | |    ||
Db       1265 QG-FQAPQD------GQG--------CVDV--------------NECELLSG----VCG 1290

Qy        268 EGF-ESPGGKITSVCTEK------------------GTWRESTLTCTEILTKINDVSL 306
              | | |:  |     || ::                |  |:      |    :|| ||
Db       1291 EAFCENVEGSFLCVCADENQEYSPMTGQCRSRATEDSGVDRQPKEEKKECYYNLNDASL 1349


Search completed: February  6, 2004, 17:38:09
Job time : 20 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2004, 17:35:09 ; Search time 21 Seconds
(without alignments)
3420.857 Million cell updates/sec

Title: US-10-066-269-58
Perfect score: 4029
Sequence: 1 MGRGPWDAGPSRRLLPLLLL.....VGLGSLAVVIILTFLSFSAV 747

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3002	74.5	570	2	T46261	hypothetical prote
2	294.5	7.3	2871	2	A55624	fibrillin-1 precur
3	290.5	7.2	3002	2	A47221	fibrillin 1 precur
4	283.5	7.0	2871	2	A55567	fibrillin I - bovi
5	283	7.0	3507	2	T34513	hypothetical prote
6	281	7.0	2907	2	A57278	fibrillin-2 precur
7	278.5	6.9	2918	2	A54105	fibrillin-2 precur
8	277.5	6.9	640	1	A30452	uromodulin precurs
9	274.5	6.8	886	2	A57172	probable hormone r
10	274	6.8	1106	2	T18739	hypothetical prote
11	272.5	6.8	1820	2	A55494	latent transformin
12	267	6.6	642	1	S52111	uromodulin precurs
13	260.5	6.5	1394	2	A35626	transforming growt
14	251	6.2	1712	2	A38261	masking protein pr
15	246.5	6.1	1247	1	MMHUND	nidogen precursor
16	244.5	6.1	705	2	S34968	fibulin, splice fo
17	243.5	6.0	685	2	S78040	fibulin, splice fo
18	241.5	6.0	2318	2	S45306	notch 3 protein -
19	240	6.0	1221	2	A49457	fibulin-2 precurso
20	239.5	5.9	2019	1	JQ1322	tenascin precursor
21	236	5.9	387	2	I38449	extracellular prot
22	236	5.9	1687	2	T30176	EGF repeat transme
23	235	5.8	644	1	A40212	uromodulin precurs
24	235	5.8	644	2	I84634	Tamm-Horsfall prot
25	234	5.8	2201	2	A32160	tenascin-C - human
26	233	5.8	570	2	A48836	fibropellin C prec
27	232	5.8	1620	2	T27283	hypothetical prote
28	231.5	5.7	2321	2	S78549	notch3 protein - h
29	231	5.7	1245	1	MMMSND	nidogen precursor
30	231	5.7	1964	2	T09059	notch4 - mouse
31	230	5.7	642	2	S53433	plasma protein S p
32	230	5.7	2524	2	A35844	Xotch protein - Af
33	229	5.7	493	2	JC5621	epidermal growth f
34	229	5.7	1133	1	EGRT	epidermal growth f
35	228	5.7	1251	2	A57293	latent transformin
36	226.5	5.6	601	2	B36346	fibulin 1 precurso
37	226	5.6	1847	2	T18308	probable vitelloge
38	225	5.6	1184	2	A55184	fibulin-2 precurso
39	224.5	5.6	683	2	C36346	fibulin 1 precurso
40	223.5	5.5	742	2	I37225	leucocyte antigen
41	222.5	5.5	835	2	JP0076	nel protein - chic
42	222	5.5	2703	1	A24420	notch protein - fr
43	221.5	5.5	2531	2	T31070	notch homolog - se
44	221	5.5	712	2	T42990	fibulin 1, splice
45	221	5.5	798	2	T22793	hypothetical prote

ALIGNMENTS

RESULT 1
T46261
hypothetical protein DKFZp761E1824.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46261
R;Bloecker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23032
A;Accession: T46261
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-570 <AAA>
A;Cross-references: EMBL:AL137432
A;Experimental source: adult amygdala; clone DKFZp761E1824
C;Genetics:
A;Note: DKFZp761E1824.1

Query Match 74.5%; Score 3002; DB 2; Length 570;
Best Local Similarity 100.0%; Pred. No. 1.4e-181;
Matches 570; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   178 DCGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEI 237
Db     1 DCGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEI 60

Qy   238 NCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEI 297
Db    61 NCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEI 120

Qy   298 LTKINDVSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPE 357
Db   121 LTKINDVSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPE 180

Qy   358 VCLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRS 417
Db   181 VCLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRS 240

Qy   418 RKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTVNVTLLRSP 477
Db   241 RKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTVNVTLLRSP 300

Qy   478 KRHSVQITIATPPAVKQTISNISGFNETCLRWRSIKTADMEEMYLFHIWGQRWYQKEFAQ 537
Db   301 KRHSVQITIATPPAVKQTISNISGFNETCLRWRSIKTADMEEMYLFHIWGQRWYQKEFAQ 360

Qy   538 EMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQITEPPLPEVEFFTVH 597
Db   361 EMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQITEPPLPEVEFFTVH 420

Qy   598 RGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNASDADGYVAAEL 657
```

```
Db        421 RGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNASDADGYVAAEL 480

Qy        658 LAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWNKVRRHSCAVWAQVKD 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LAKDVPDDAMEIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWNKVRRHSCAVWAQVKD 540

Qy        718 SSLMLLQMAGVGLGSLAVVIILTFLSFSAV 747
              |||||||||||||||||||||||||||||
Db        541 SSLMLLQMAGVGLGSLAVVIILTFLSFSAV 570


RESULT 2
A55624
fibrillin-1 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 02-Aug-2002
C;Accession: A55624
R;Yin, W.; Smiley, E.; Germiller, J.; Sanguineti, C.; Lawton, T.; Pereira, L.; Ramirez,
J. Biol. Chem. 270, 1798-1806, 1995
A;Title: Primary structure and developmental expression of Fbn-1, the mouse fibrillin ge
A;Reference number: A55624; MUID:95130561; PMID:7829516
A;Accession: A55624
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2871 <YIN>
A;Cross-references: GB:L29454; NID:g575509; PIDN:AAA56840.1; PID:g575510
C;Genetics:
A;Gene: Fbn-1
C;Superfamily: fibrillin 1; EGF homology
F;1201-1236/Domain: EGF homology <EGF>

  Query Match             7.3%;  Score 294.5;  DB 2;  Length 2871;
  Best Local Similarity   26.7%;  Pred. No. 3.4e-10;
  Matches   83;  Conservative   33;  Mismatches   94;  Indels  101;  Gaps   15;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  |:| | | |||: ||   || |  | ||
Db       1289 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCGRHAVCTNTA 1343

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
              | | | |  |:           ||  |||:|||   : :|     | || ||: | | |
Db       1344 GSFKCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKD 1393

Qy        155 GYLPRNGPEPFHPTTDATSCTEID--------CGTPP--EVPDGYIIGNYTSSLGSQVRY 204
              ||           | |  :||::|        ||       | ||              |
Db       1394 GY----------TGDGFTCTDLDECSENLNLCGNGQCLNAPGGY------------RC 1429

Qy        205 ACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARY 264
              |  ||  ||     :|  :     | : |     | | |                 |: |
Db       1430 ECDMGF--VPSADGKACEDIDECSLPNI-CVFGTCHNLP---------------GLFRC 1470

Qy        265 VCQEGFE--SPGGKITSVCTEKGTWRESTLPCTEILTKINDVSLFNDTCVR--------- 313
              |: |:|    ||                ||::   ::  :  :  ||
Db       1471 ECEIGYELDRSGG----------------NCTDVNECLDPTTCISGNCVNTPGSYTCDC 1513

Qy        314 ---WQINSRRI 321
                 :::|  |:
Db       1514 SPDFELNPTRV 1524


RESULT 3
A47221
fibrillin 1 precursor - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1995 #sequence_revision 25-Apr-1997 #text_change 02-Aug-2002
C;Accession: A47221; I54355; S17064; I59574; S17062; S62111; A34198
R;Corson, G.M.; Chalberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y.
Genomics 17, 476-484, 1993
A;Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structu
A;Reference number: A47221; MUID:94010947; PMID:7691719
A;Accession: A47221
A;Molecule type: mRNA
A;Residues: 1-337,'T',339-1029 <COR>
A;Cross-references: GB:X63556
R;Pereira, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Pangilinan, T.; Bona
Hum. Mol. Genet. 2, 961-968, 1993
A;Title: Genomic organization of the sequence coding for fibrillin, the defective gene
A;Reference number: I54355; MUID:93372860; PMID:8364578
A;Accession: I54355
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 132-3002 <PER>
A;Cross-references: GB:L13923; NID:g306745; PIDN:AAB02036.1; PID:g306746
R;Maslen, C.L.; Corson, G.M.; Maddox, B.K.; Glanville, R.W.; Sakai, L.Y.
Nature 352, 334-337, 1991
A;Title: Partial sequence of a candidate gene for the Marfan syndrome.
A;Reference number: S17064; MUID:91304568; PMID:1852207
A;Accession: S17064
A;Molecule type: mRNA
A;Residues: 1030-3002 <MAS>
A;Cross-references: EMBL:X63556
R;Dietz, H.C.; Valle, D.; Francomano, C.A.; Kendzior, R.J.
Science 259, 680-683, 1993
A;Title: The skipping of constitutive exons in vivo induced by nonsense mutations.
A;Reference number: I59574; MUID:93157831; PMID:8430317
A;Accession: I59574
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 2217-2288,'I',2290-2325 <RES>
A;Cross-references: GB:S54426; NID:g264860; PIDN:AAB25244.1; PID:g264861
R;Lee, B.; Godfrey, M.; Vitale, E.; Hori, H.; Mattei, M.G.; Sarfarazi, M.; Tsipouras, P
Nature 352, 330-334, 1991
A;Title: Linkage of Marfan syndrome and a phenotypically related disorder to two differ
A;Reference number: S17062; MUID:91304567; PMID:1852206
A;Accession: S17062
A;Molecule type: mRNA
A;Residues: 'VLVTVVFIFLSYNKML',944-1444 <LEE1>
A;Cross-references: EMBL:X62008; NID:g31398; PIDN:CAB56534.1; PID:g5924015
A;Accession: S62111
A;Molecule type: protein
A;Residues: 1166-1176,'X',1178-1180,'D',1182-1185 <LEE2>
R;Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.
J. Biol. Chem. 264, 21381-21385, 1989
A;Title: Connective tissue microfibrils. Isolation and characterization of three large
A;Reference number: A34198; MUID:90078246; PMID:2512293
A;Accession: A34198
A;Molecule type: protein
A;Residues: 565-575;1890-1892,'I',1894-1900 <MAD>
C;Comment: Fibrillin is a major component of elastin-associated microfibrils.
C;Genetics:
A;Gene: GDB:FBN1
A;Cross-references: GDB:127115; OMIM:134797; OMIM:154700
A;Map position: 15q21.1-15q21.1
A;Introns: 2236/1; 2258/1; 2297/1
C;Superfamily: fibrillin 1; EGF homology
C;Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein;
F;1-3002/Product: fibrillin (5'-region exon A splice form) (fragment) #status predicted
F;132-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MATC>
F;1332-1367/Domain: EGF homology <EGF>
F;1457-1492/Domain: EGF homology <EGF2>
F;2262-2295/Domain: EGF homology <EGF1>

  Query Match             7.2%;  Score 290.5;  DB 2;  Length 3002;
  Best Local Similarity   26.4%;  Pred. No. 6.5e-10;
  Matches   82;  Conservative   34;  Mismatches   94;  Indels  101;  Gaps   15;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  |:| | | |||: ||   || |  | ||
Db       1420 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCGKHAVCTNTA 1474

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
              | | | |  |:           ||  |||:|||   : :|     | || ||: | | :
Db       1475 GSFKCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKE 1524

Qy        155 GYLPRNGPEPFHPTTDATSCTEID--------CGTPP--EVPDGYIIGNYTSSLGSQVRY 204
              ||           | |  :||::|        ||       | ||              |
```

```
Db       1525 GY-----------TGDGFTCTDLDECSENLNLCGNGQCLNAPGGY------------RC 1560

Qy        205 ACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARY 264
              |  ||  ||     :|  :     | : |    | | |                |: |
Db       1561 ECDMGF--VPSADGKACEDIDECSLPNI-CVFGTCHNLP---------------GLFRC 1601

Qy        265 VCQEGFE--SPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVR--------- 313
              |: |:|    ||                ||::   ::  :  :  ||
Db       1602 ECEIGYELDRSGG----------------NCTDVNECLDPTTCISGNCVNTPGSYICDC 1644

Qy        314 ---WQINSRRI 321
                 :::|  |:
Db       1645 PPDFELNPTRV 1655


RESULT 4
A55567
fibrillin I - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
C;Accession: A55567
R;Tilstra, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
Genomics 23, 480-485, 1994
A;Title: Sequence of the coding region of the bovine fibrillin cDNA and localization to
A;Reference number: A55567; MUID:95137597; PMID:7835900
A;Accession: A55567
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2871 <TIL>
A;Cross-references: GB:L28748; NID:g508427; PIDN:AAA74122.1; PID:g508428
C;Superfamily: fibrillin 1; EGF homology
F;1201-1236/Domain: EGF homology <EGF>

  Query Match             7.0%;  Score 283.5;  DB 2;  Length 2871;
  Best Local Similarity   26.0%;  Pred. No. 1.7e-09;
  Matches   81;  Conservative   34;  Mismatches   95;  Indels  101;  Gaps   15;

Qy         37 DVCATCHEHATCQQREGKKICICNYGFVG-NGRTQCVDKNECQFGATLVCGNHTSCHNTP 95
              ::| :     ||:  :|  || |: |: |  |:| | | |||: ||   |  |  | ||
Db       1289 NICLS----GTCENTKGSFICHCDMGYSGKKGKTGCTDINECEIGAH-NCDRHAVCTNTA 1343

Qy         96 GGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCVNTHGSFECYCMD 154
              | | | |  |:             ||  |||:|||   : :|     | || ||: | | :
Db       1344 GSFKCSCSPGW----------IGDGIKCTDLDECSNGTHMCSQHADCKNTMGSYRCLCKE 1393

Qy        155 GYLPRNGPEPFHPTTDATSCTEID--------CGTPP--EVPDGYIIGNYTSSLGSQVRY 204
              ||           | |  :||::|        ||      | ||             |
Db       1394 GY-----------TGDGFTCTDLDECSENLNLCGNGQCLNAPGGY------------RC 1429

Qy        205 ACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARY 264
              |  ||  ||     :|  :     | : |    | | |                |: |
Db       1430 ECDMGF--VPSADGKACEDIDECSLPNI-CVFGTCHNLP---------------GLFRC 1470

Qy        265 VCQEGFE--SPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVR--------- 313
              |: |:|    ||                ||::   ::  :  :  ||
Db       1471 ECEIGYELDRSGG----------------NCTDVNECLDPTTCISGNCVNTPGSYTCDC 1513

Qy        314 ---WQINSRRI 321
                 :::|  |:
Db       1514 PPDFELNPTRV 1524


RESULT 5
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34513
R;Favello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of C. elegans cosmid ZK783.
A;Reference number: Z21536
A;Accession: T34513
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3507 <FAV>
A;Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A;Experimental source: strain Bristol N2; clone ZK783
C;Genetics:
A;Gene: CESP:ZK783.1
A;Map position: 3
A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
 3504/1

  Query Match             7.0%;  Score 283;  DB 2;  Length 3507;
  Best Local Similarity   21.8%;  Pred. No. 2.3e-09;
  Matches  185;  Conservative   81;  Mismatches  303;  Indels  280;  Gaps   40;

Qy         22 GLAR---GAAGAPGPDG---LDVCAT----CHEHATCQQREGKKICICNYGFVGNGRTQC 71
              |: |   |  |||       :| |||    ||| | ||  |  | | |  ||       |
Db       1431 GVCRCEPGFEGAPPKKSCVDVDECATGDHNCHESARCQNYVGGYACFCPTGFRKADDGSC 1490

Qy         72 VDKNECQFGATLVCGNHTSCHNTPG----------GFYCI---------------- 101
              | :||    :  || :  | | ||           |: |:
Db       1491 QDIDECTEHNSTCCGANAKCVNKPGTYSCECENGFLGDGYQCVPTTKKPCDSTQSSKSHC 1550

Qy        102 --------------------CLEGYR----------------------ATNNNKTF-- 115
                                  |: ||:                      | | ||
Db       1551 SESNMSCEVDTVDGSVECKECMGGYKKSGKVCEDINECVAEKAPCSLNANCVNMNGTFSC 1610

Qy        116 -----IPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTD 170
                  ||  ||||:||:     |     | |  |||:| |  |:      :  :|
Db       1611 SCKQGYRGDGFMCTDINECDERHPCHPHAECTNLEGSFKCECHSGFEGDGIKKCTNPL-- 1668

Qy        171 ATSCTEID--CGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWE 228
              || :::  ||     |     :  | || |     | ||    |   :|| :   |
Db       1669 ERSCEDVEKFCGRVDHV-SCLSVRIYNGSLSSVCE--CEPGFRFEKES--NSCVDIDECE 1723

Qy        229 SPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWR 288
              : :|  :          |: |       | | | ||:|  ||    |||:
Db       1724 ESRNNCDPAS---------AVCVNTE-----GSYRCECAEGYEGEGG----VCTDIDECD 1765

Qy        289 ESTLTCTEILTKINDVSLF----------NDTCVRWQINSRRINPKISYVISIKGQRLD 337
              |  :   || :                ||:: :       ||     | |     |
Db       1766 RGMAGCDSMAMCINRMGSCGCKCMAGYTGDGATCIKIE-----EEPK-----SDKTACTD 1815

Qy        338 PMESVRE-ETVNLTTDSRTPEVCLALYPGTNYTVN----------------------I 372
              : | |    | |       | |  || :: :|                      |
Db       1816 EWSRLCELEKKQCTVDEEEVPQCGACLPG-HHPINGTCQSLQISGLCAQKNDCNKHAECI 1874

Qy        373 STAPPRRSMPAVIGF----------QTAEVDLLEDD--------GSFN------ISIFN 407
                |          ||          :    : :|:          ||||      :
Db       1875 DIHPDSHFCSCPDGFIGDGMICDDVDECNNAGMCDDENTKCENTIGSFNCVCLEGFKKVD 1934

Qy        408 ETCL-----KLNRRSRKVGSEH-------MYQFTVLGQRWYLANFSHATSFNFTTREQVP 455
              | |:     : ||   ::  |:         : |  |    : : : |||   || |  |
Db       1935 EKCVVDEKKQPNREKIEIDEENSSSSNSGQEKPTTKG----IVSSTSATSSESTTAE--P 1988

Qy        456 VVCLDLYPTTDYTVNVTLLRSPKR----------HSVQITIATPPAVKQTISNIS--- 500
              |   : ||  | ::|  :||:              | : | |:   |  | |  |
Db       1989 HVTTSISSTTS-TKDMTSSKSPENVTMSSESPEVSTSSSKSTTASETTVSSTPSESSSSE 2047

Qy        501 -------GFNETCLRWRSIKTADMEEMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEV- 552
                    :   |:|:   :|                | : |:|  :|| |  |||
Db       2048 APLTSSPATTTEVITESSVKSTTPKE----------ESSSEITVKLSSKS--PEVT 2091

Qy        553 --CLDLRPGTNYNVS----------LRALSSELPVVISLTTQI-----TEPPLP---- 589
              :  | |    |                |||| ||  :  |::     |:| |
Db       2092 ESSVKSSPSTPSTTSQSVTSTVPETSKSTVLSSEAPVTSTSPTEVHTSSETKPSLSASST 2151

Qy        590 --EVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNAS 647
              :    |    | ::  | | |  :|    | |: | |: | | |    |:   :|:
Db       2152 TGDTNSTTPSTSSLASVKSTSAPE--GTSAS----VAPVKL-SSLSPDVSQPSTKTFDAT 2204

Qy        648 DADGYVAAE 656
                ::    |:|
```

```
Db       2205 ESSTVQASE 2213
```

RESULT 6
A57278
fibrillin-2 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 02-Aug-2002
C;Accession: A57278
R;Zhang, H.; Hu, W.; Ramirez, F.
J. Cell Biol. 129, 1165-1176, 1995
A;Title: Developmental expression of fibrillin genes suggests heterogeneity of extracell
A;Reference number: A57278; MUID:95263670; PMID:7744963
A;Accession: A57278
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2907 <ZHA>
A;Cross-references: GB:L39790; NID:g762830; PIDN:AAA74908.1; PID:g762831
C;Superfamily: fibrillin 1; EGF homology
F;1239-1274/Domain: EGF homology <EGF1>
F;2488-2523/Domain: EGF homology <EGF>

Query Match 7.0%; Score 281; DB 2; Length 2907;
Best Local Similarity 28.8%; Pred. No. 2.5e-09;
Matches 83; Conservative 39; Mismatches 114; Indels 52; Gaps 15;

```
Qy         25 RGAAGAPGPDGLDVCA-TCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGATL 83
Db       1353 KGTTGCTDVDECEIGAHNCDMHASCLNVPGSFKCSCREGWVGNG-IKCIDLDECANG-TH 1410

Qy         84 VCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLCRHGGRCV 142
Db       1411 QCSINAQCVNTPGSYRCACSEGF----------TGDGFTCSDVDECAENTNLCEN-GQCL 1459

Qy        143 NTHGSFECYCMDGYLPRNGPEPFHPTTDATRCTEID-CGTPPEVPDGYIIGNYTSSLGSQ 201
Db       1460 NVPGAYRCECEMG---------FTPASDSRRCQDIDECSF-----QNICVFGTCNNLPGM 1505

Qy        202 VRYACREGFFSVPEDTVSSCTGLGTWESPKLHC----------QEINCGNPPEMR-HAI 249
Db       1506 FHCICDDGY--ELDRTGGNCTDIDECADP-TNCVNGLCVNTPGRYECNC--PPDFQLNAT 1560

Qy        250 LVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEI 297
Db       1561 GVGCVDNRVGN-----CYLKFGPRGDGSLSCNTEAGVGVSRSSCCCSL 1603
```

RESULT 7
A54105
fibrillin-2 precursor - human
C;Species: Homo sapiens (man)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 02-Aug-2002
C;Accession: A54105; S17063; S31101
R;Zhang, H.; Apfelroth, S.D.; Hu, W.; Davis, E.C.; Sanguineti, C.; Bonadio, J.; Mecham,
J. Cell Biol. 124, 855-863, 1994
A;Title: Structure and expression of fibrillin-2, a novel microfibrillar component prefe
A;Reference number: A54105; MUID:94165150; PMID:8120105
A;Accession: A54105
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A;Molecule type: mRNA
A;Residues: 1-2918 <ZHA>
A;Cross-references: GB:U03272
R;Lee, B.; Godfrey, M.; Vitale, E.; Hori, H.; Mattei, M.G.; Sarfarazi, M.; Tsipouras, P.
Nature 352, 330-334, 1991
A;Title: Linkage of Marfan syndrome and a phenotypically related disorder to two differe
A;Reference number: S17062; MUID:91304567; PMID:1852206
A;Accession: S17063
A;Molecule type: mRNA
A;Residues: 752-1489,1791,'GS',1794-1796,'QLI',1922-1923,'LD',1926,'P',1928 <LEE>
A;Cross-references: EMBL:X62009
R;Milewicz, D.M.
submitted to the EMBL Data Library, December 1992
A;Reference number: S31101
A;Accession: S31101
A;Molecule type: mRNA
A;Residues: 752-1407,'R',1409-1489,1791,'GS',1794-1796,'QLI',1922-1923,'LD',1926,'P',19
A;Cross-references: EMBL:X62009
C;Genetics:
A;Gene: GDB:FBN2
A;Cross-references: GDB:128122; OMIM:121050
A;Map position: 5q23-5q31
C;Superfamily: fibrillin 1; EGF homology
C;Keywords: extracellular protein
F;1-29/Domain: signal sequence #status predicted <SIG>
F;30-2918/Product: fibrillin-2 #status predicted <MAT>
F;1245-1280/Domain: EGF homology <EGF1>
F;1970-2013/Domain: EGF homology <EGF>

Query Match 6.9%; Score 278.5; DB 2; Length 2918;
Best Local Similarity 28.8%; Pred. No. 3.6e-09;
Matches 90; Conservative 30; Mismatches 126; Indels 67; Gaps 17;

```
Qy          2 GRGPWDAGPSRRLLPLLLLLGLARG------AAGAPGP--DGLDV----CATCHEHAT-- 47
Db        449 GYGPGGTG----FIPIPGGNGFSPGVGGAGVGAGGQGPIITGLTILNQTIDICKHHANLC 504

Qy         48 ----CQQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICL 103
Db        505 LNGRCIPTVSSYRCECNMGYKQDANGDCIDVDEC---TSNPCTN-GDCVNTPGSYYCKCH 560

Qy        104 EGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPE 163
Db        561 AGFQRTPTKQA--------CIDIDECIQNGVLCKNGRCVNSDGSFQCICNAG-------- 604

Qy        164 PFHPTTDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCT 222
Db        605 -FELTTDGKNCVDHDECTTTNMCLNGMCINE-----DGSFKCICKPGFVLAPNGRY--CT 656

Qy        223 GLGTWESPKL----HC----QEINCGNPP----EMRHAILVGNH--SSRLGGVARYVCQE 268
Db        657 DVDECQTPGICMNGHCINSEGSFRCDCPPGLAVGMDGRVCVDTHMRSTCYGGIKKGVCVR 716

Qy        269 GFESPGGKITSVC 281
Db        717 PF--PGAVTKSEC 727
```

RESULT 8
A30452
uromodulin precursor - human
N;Alternate names: Tamm-Horsfall urinary glycoprotein; uromucoid
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A30452; A30453
R;Pennica, D.; Kohr, W.J.; Kuang, W.J.; Glaister, D.; Aggarwal, B.B.; Chen, E.Y.; Goedd
Science 236, 83-88, 1987
A;Title: Identification of human uromodulin as the Tamm-Horsfall urinary glycoprotein.
A;Reference number: A30452; MUID:87177970; PMID:3453112
A;Accession: A30452
A;Molecule type: DNA; mRNA
A;Residues: 1-640 <PEN>
A;Cross-references: GB:M15881; NID:g340163; PIDN:AAA36798.1; PID:g340164
A;Note: parts of this sequence, including the amino end of the mature protein, were det
R;Hesion, C.; Decker, J.M.; Sherblom, A.P.; Kumar, S.; Yue, C.C.; Mattaliano, R.J.; Tiz
Science 237, 1479-1484, 1987
A;Title: Uromodulin (Tamm-Horsfall glycoprotein): A renal ligand for lymphokines.
A;Reference number: A30453; MUID:87319675; PMID:3498215
A;Accession: A30453
A;Molecule type: mRNA
A;Residues: 1-640 <HES>
A;Cross-references: GB:M17778
A;Note: the authors translated the codon GTG for residue 381 as Asp
A;Note: parts of this sequence, including the amino end of the mature protein, were det
R;Rindler, M.J.; Naik, S.S.; Li, N.; Hoops, T.C.; Peraldi, M.N.
J. Biol. Chem. 265, 20784-20789, 1990
A;Title: Uromodulin (Tamm-Horsfall Glycoprotein/Uromucoid) is a phosphatidylinositol-li
```
```

```
A;Reference number: A30454; MUID:91065873; PMID:2249987
A;Contents: annotation; GPI-anchor
C;Comment: The protein is anchored to the cell membrane by a phosphatidylinositol linkag
e membrane-bound form.
C;Genetics:
A;Gene: GDB:UMOD
A;Cross-references: GDB:133729; OMIM:191845
A;Map position: 16p13.11-16p12.3
A;Introns: 30/1; 289/1; 325/1; 394/3; 444/2; 526/2; 580/3; 608/1; 621/1
A;Note: the first intron occurs before the initiator codon
C;Superfamily: uromodulin; EGF homology; membrane glycoprotein 2 homology; ZP domain hom
C;Keywords: duplication; glycoprotein; lipoprotein; phosphatidylinositol linkage; phosph
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-640/Product: uromodulin #status predicted <MAT>
F;69-106/Domain: EGF homology <EGF>
F;170-639/Domain: membrane glycoprotein 2 homology <MGH>
F;334-583/Domain: ZP domain homology <ZPH>
F;625-640/Domain: transmembrane #status predicted <TRM>
F;38,76,80,275,322,513/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;232,396/Binding site: carbohydrate (Asn) (covalent) #status experimental

  Query Match             6.9%;  Score 277.5;  DB 1;  Length 640;
  Best Local Similarity   23.4%;  Pred. No. 6.5e-10;
  Matches  163;  Conservative   77;  Mismatches  217;  Indels  239;  Gaps   38;

Qy         39 CATCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGF 98
              |: || :||| : |    | |  || |:| | ||| :||       |  ::|| |||| |
Db         32 CSECHSNATCTEDEAVTTCTCQEGFTGDGLT-CVDLDECAIPGAHNCSANSSCVNTPGSF 90

Qy         99 YCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGL--CRHGGRCVNTHGSFECYCMDGY 156
               |:| ||:|          : |  |||:|||   ||  |     |||  ||: | |  ||
Db         91 SCVCPEGFRL---------SPGLGCTDVDECAEPGLSHCHALATCVNVVGSYLCVCPAGY 141

Qy        157 --------------------LPRN------GPEPFHPTTD-------------------- 170
                              :|       |   | | |
Db        142 RGDGWHCECSPGSCGPGLDCVPEGDALVCADPCQAHRTLDEYWRSTEYGEGYACDTDLRG 201

Qy        171 ------------ATSCTEI-DC---------GTPPEVPDGYIIGNYTS---------SLG 199
                         | :|  :  |          ||| |   :| :     :
Db        202 WYRFVGQGGARMAETCVPVLRCNTAAPMWLNGTHPSSDEGIVSRKACAHWSGHCCLWDAS 261

Qy        200 SQVRYACREGFF-----SVPEDTVSSCTG----LGTWE-----------SPKLHC---QE 236
               ||: ||  |::     : ||  :: ||      || |           : : ||   |:
Db        262 VQVK-ACAGGYYVYNLTAPPECHLAYCTDPSSVEGTCEECSIDEDCKSNNGRWHCQCKQD 320

Qy        237 INCGNPPEMRHAILVGNHSSRLGGVARYVCQ---EGFESPGGKI-----TSVCT-----E 283
              |  :   : | :  |  :  :   |:   ||    ||:    |:       | |:     :
Db        321 FNITDISLLEHRLECGANDMK---VSLGKCQLKSLGFD----KVFMYLSDSRCSGFNDRD 373

Qy        284 KGTWRESTLT------CTEILTKINDVSLFNDTC-VRWQINSRRINPKISYV-------- 328
                | | :|       |  :||:    : :::|  :  :|  | :| ||::
Db        374 NRDW-VSVVTPARDGPCGTVLTRNETHATYSNTLYLADEIIIRDLNIKINFACSYPLDMK 432

Qy        329 ISIKGQRLDPMES---VREETVNLTTDSRTPEVCLALYPGTNYTVNISTAPPRRSMPAVI 385
              :|:|   | || |   :|     : |       | :||:   :||            |
Db        433 VSLK-TALQPMVSALNIRVGGTGMFT------VRMALFQTPSYT-----------QP--- 471

Qy        386 GFQTAEVDLLEDDGSFNISIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATS 445
              :| : | |      |    ::  | |      ||          | :|    |    |:||
Db        472 -YQGSSVTL-----STEAFLYVGTMLDGGDLSR---------FALLMTNCYATPSSNATD 516

Qy        446 --FNFTTREQVPVVCLDLYPTTDYTVNVTL--LRSPKRHSVQITIATPPAVKQTISNISG 501
                |  ::: |        | | |: |       |  | |||:
Db        517 PLKYFIIQDRCP-------HTRDSTIQVVENGESSQGRFSVQM---------------- 552

Qy        502 FNETCLRWRSIKTADME----EMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLR 557
                    :|     |:     |:||     :|       |  | |  |     |:|
Db        553 -------FRFAGNYDLVYLHCEVYLCDTMNEKC--KPTCSGTRFR-SGSVIDQSRVLNLG 602

Qy        558 PGTNYNV---------SLRALSSELPVVISLTTQIT 584
              | |   |         ||  |   ||:::| |  :|
Db        603 PITRKGVQATVSRAFSSLGLLKVWLPLLLSATLTLT 638
```

```
RESULT 9
A57172
probable hormone receptor EMR1 precursor - human
C;Species: Homo sapiens (man)
C;Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 11-Jan-2000
C;Accession: A57172
R;Baud, V.; Chissoe, S.L.; Viegas-Pequignot, E.; Diriong, S.; N'Guyen, V.C.; Roe, B.A.;
Genomics 26, 334-344, 1995
A;Title: EMR1, an unusual member in the family of hormone receptors with seven transmemb
A;Reference number: A57172; MUID:95324926; PMID:7601460
A;Accession: A57172
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-886 <BAU>
A;Cross-references: GB:X81479; NID:g784993; PIDN:CAA57232.1; PID:g784994
C;Genetics:
A;Gene: GDB:EMR1
A;Cross-references: GDB:378349; OMIM:600493
A;Map position: 19p13.3-19p13.3
C;Superfamily: unassigned EGF-related proteins; EGF homology
C;Keywords: transmembrane protein
F;136-170/Domain: EGF homology <EGF>

  Query Match             6.8%;  Score 274.5;  DB 2;  Length 886;
  Best Local Similarity   22.8%;  Pred. No. 1.5e-09;
  Matches   89;  Conservative   53;  Mismatches  125;  Indels  123;  Gaps   17;

Qy         42 CHEHATCQQREGKKICICNYGFV-GNGR-------TQCVDKNECQFGATLVCGNHTSCHN 93
                |  :|||         | |  ||:  ||:        :| | :||   :   || ::|| |
Db         41 CPAYATCTNTVDSYYCTCKQGFLSSNGQNHFKDPGVRCKDIDECS-QSPQPCGPNSSCKN 99

Qy         94 TPGGFYCICLEGYRATNNNKTFIPNDGTF-CTDIDECEVSGLCRHGGRCVNTHGSFECYC 152
                 | : | ||:|: :   |       | | ||||:||  | :|     |||: ||: | |
Db        100 LSGRYKCSCLDGFSSPTGNDWVPGKPGNFSCTDINECLTSRVCPEHSDCVNSMGSYSCSC 159

Qy        153 MDGYLPRNGPEPFHPTTDATSCTEI-DCGTPPEVPD----GYIIGNYTSSLGSQVRYACR 207
                |:: ||          ::| :: :|  |    |:       :|||:          |
Db        160 QVGFISRN----------STCEDVNECADPRACPEHATCNNTVGNYSC--------FCN 200

Qy        208 EGFFSVPEDTVSSCTGLGTWESPKLHCQEIN----------CGNPPEMRHAILVGNHSS 256
              || |        || ||      |  |::|:                |  |  |
Db        201 PGFES--SSGHLSCQGL------KASCEDIDECTEMCPINSTCTNTP------------ 239

Qy        257 RLGGVARYVCQEGFESPGGKI----------------------TSVCTE--------- 283
                |      |  ||     |::                        |:||
Db        240 ---GSYFCTCHPGFAPSSGQLNFTDQGVECRDIDECRQDPSTCGPNSICTNALGSYSCGC 296

Qy        284 --------KGTWRESTLTCTEILTKINDVSLFNDTCVRWQINSRRINP-------KISYV 328
                      :|: ::   :|  :| |  :  : ::  ::       : |       :|: :
Db        297 IVGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNKQIQQCQEGTAVKPAYVSFCAQINNI 356

Qy        329 ISIKGQRLDPMESVREETVNL--TTDSRTP 356
                |:    || :   :    |:|  ||:|  |
Db        357 FSV----LDKVCENKTTVVSLKNTTESFVP 382
```

```
RESULT 10
T18739
hypothetical protein B0393.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18739
R;Sulston, J.
submitted to the EMBL Data Library, September 1994
A;Reference number: Z19013
A;Accession: T18739
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1106 <WIL>
A;Cross-references: EMBL:Z37983; NID:e1519039; PIDN:CAA86058.1; GSPDB:GN00021; CESP:B039
A;Experimental source: clone B0393
C;Genetics:
```

```
A;Gene: CESP:B0393.5
A;Map position: 3
A;Introns: 15/3; 60/1; 100/3; 343/2; 443/1; 507/3; 609/3; 659/1; 796/1; 876/1; 946/1; 99

  Query Match            6.8%;  Score 274;  DB 2;  Length 1106;
  Best Local Similarity  25.0%;  Pred. No. 2.1e-09;
  Matches  107;  Conservative   53;  Mismatches  144;  Indels  124;  Gaps   24;

Qy         39 CATCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQF-GATLVCGNHTSCHNTPGG 97
              || | ::| |        :| |: || |:|  :| | :||:  ||  || :|: | || |
Db        373 CARCDQNAKC----SNGVCTCSEGFTGDG-FRCYDVDECEIPGA--VCRDHSICSNTIGS 425

Qy         98 FYCICLEGYRATNNNKTFIPNDGTFCTDIDEC-EVSGLC---RHGGRCVNTHGSFECYCM 153
              | | |  |||          ||  | |:|| |:  :|     | :|:|  |:||| |
Db        426 FECTCHGGYRF---------EDGK-CEDVDECRELPKICGDPNKGTKCINKDGTFECLCK 475

Qy        154 DGYLPRNGPEPFHPTTDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFS 212
              |||      :|     |   | || :  :    | |           |  ||
Db        476 DGY----EGDPSSECRDVNECKNPDACGPNSQCTNTQ--GGYECE--------CLAGFER 521

Qy        213 VPEDTVSSCTGLGTWESPKLHCQEIN-CGNPPEMRHAILVGNHSSRLGGVARYVCQEGFE 271
              : |                  || : : |   |    ||      |   |  :  |::||
Db        522 IAEGA--------------HCTDRDECAVEPCHPAAI-----CSNTRGSYKCECRDGFV 561

Qy        272 SPGGKITSVCTE------------KGTWRESTLTCTEILTKINDVSLFNDTCVRWQINS 318
              |    | |              :|  ||     :  ::::::|    :  :::
Db        562 GDG----KTCHETILYPISNDSTVIPRSWDSST-----AVPILSEITIFGKKYSKIYLST 612

Qy        319 RRINPKISYVISIKGQRLDPMES----------------VREETVNLTTDSRTPEVCLALY 363
              |   ||:  ::|  :|  |:              :|| |  |  : |  | | |
Db        613 NGI---ISFDGPLQG-LIDHAETLKKPAIFALHAQFDYIREGLVAYTFINDTDAVTLPLL 668

Qy        364 PGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRSRKVGSE 423
              :: | ||       :  |||  :  |                    | |: |||
Db        669 --MRSSIGIQTA------MGIENFQTKRLHIFTFD----------------RVRQSGSE 703

Qy        424 HMYQFTVL 431
              :  | ::
Db        704 NFNSFQIV 711


RESULT 11
A55494
latent transforming growth factor-beta-binding protein - human
C;Species: Homo sapiens (man)
C;Date: 03-Mar-1995 #sequence_revision 03-Mar-1995 #text_change 11-Jan-2000
C;Accession: A55494
R;Moren, A.; Olofsson, A.; Stenman, G.; Sahlin, P.; Kanzaki, T.; Claesson-Welsh, L.; ten
J. Biol. Chem. 269, 32469-32478, 1994
A;Title: Identification and characterization of LTBP-2, a novel latent transforming grow
A;Reference number: A55494; MUID:95096101; PMID:7798248
A;Accession: A55494
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1820 <MOR>
A;Cross-references: GB:Z37976
C;Genetics:
A;Gene: GDB:LTBP2
A;Cross-references: GDB:568901
A;Map position: 11pter-11qter
C;Superfamily: unassigned EGF-related proteins; EGF homology
F;1222-1257/Domain: EGF homology <EGF>
F;1525-1565/Domain: EGF homology <EGF1>

  Query Match            6.8%;  Score 272.5;  DB 2;  Length 1820;
  Best Local Similarity  28.1%;  Pred. No. 4.8e-09;
  Matches   82;  Conservative   25;  Mismatches   90;  Indels   95;  Gaps   16;

Qy         22 GLARGAAGA----------PGPDGLDVCATCHE---------HATCQQREGKKICICNY 61
              |:   ||:           | | | | |         |:    |   |:|
Db       1106 GVCTNTAGSFSCKDCDGGYRPSPLG-DSCEDVDECEDPQSSCLGGECKNTVGSYQCLCPQ 1164

Qy         62 GF-VGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDG 120
```

```
Db       1165 GFQLANG-TVCEDVNECM--GEEHCAPHGECLNSHGSFPCLCAPGFVSAEG--------G 1213

Qy        121 TFCTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEID-- 178
              | | |:|||  :  |  || |||| ||| | |  |          | |: ::  | :||
Db       1214 TSCQDVDECATTDPC-VGGHCVNTEGSFNCLCETG---------FQPSPESGECVDIDEC 1263

Qy        179 -------CGT-----PPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGT 226
                    |||     |        |:|   ||      |: ||   |
Db       1264 EDYGDPVCGTWKCENSP--------GSYRCVLG------CQPGFHMAPNG---------- 1299

Qy        227 WESPKLHCQEINCGNPPEMRHAILVGNHS--SRLGGVARYVCQEGFE-SPGG 275
                    :| :  |  :  : |:|       |  | :| :||| || |
Db       1300 -----------DCIDIDECANDTMCGSHGFCDNTDGSFRCLCDQGFEISPSG 1340


RESULT 12
S52111
uromodulin precursor - mouse
N;Alternate names: Tamm-Horsfall urinary glycoprotein; uromucoid
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S52111
R;Prasadan, K.; Bates, J.; Badgett, A.; Dell, M.; Sukhatme, V.; Yu, H.; Kumar, S.
Biochim. Biophys. Acta 1260, 328-332, 1995
A;Title: Nucleotide sequence and peptide motifs of mouse uromodulin (Tamm-Horsfall prot
A;Reference number: S52111; MUID:95178555; PMID:7873609
A;Accession: S52111
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-642 <PRA>
C;Superfamily: uromodulin; EGF homology; membrane glycoprotein 2 homology; ZP domain ho
C;Keywords: duplication; glycoprotein; lipoprotein; membrane protein; phosphatidylinosi
F;69-105/Domain: EGF homology <EG1>
F;111-147/Domain: EGF homology <EG2>
F;171-641/Domain: membrane glycoprotein 2 homology <MGH>
F;335-584/Domain: ZP domain homology <ZPH>

  Query Match            6.6%;  Score 267;  DB 1;  Length 642;
  Best Local Similarity  24.2%;  Pred. No. 3e-09;
  Matches   99;  Conservative   38;  Mismatches  140;  Indels  132;  Gaps   16;

Qy         28 AGAPGPDGLDVCATCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGN 87
              |||         |: || :|||        | |  || |:|   | | :||    |  | |
Db         21 AGASNSTEARRCSECHNNATCTVDGVVTTCSCQTGFTGDGLV-CEDMDECATPWTHNCSN 79

Qy         88 HTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGL--CRHGGRCVNTH 145
               :|| |||| | | | :|:| |           |  |||:|||   ||  |     ||||
Db         80 -SSCVNTPGSFKCSCQDGFRLT---------PGLGCTDVDECSEQGLSNCHALATCVNTE 129

Qy        146 GSFECYCMDGYLPRNGPEPFHPTTDATSCTE-IDCGTPPEVPDGYII----------GN 193
              | : | |  |:      : ::     :||   :||   |: ||| ::
Db        130 GDYLCVCPKGFT----GDGWYCECSPSSCEPGLDC--LPQGPDGKLVCQDPCNTYETLTE 183

Qy        194 YTSSLGSQVRYAC---REGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAIL 250
              |  |    | |:|   : |::       || |    :    : |   |    |
Db        184 YWRSTEYGVGYSCDAGQHGWYRF--------TGQGGVRMAETCVPVLACNTAAPM---WL 232

Qy        251 VGNH-SSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFND 309
              |:| ||  | |:|  |                                    ::|
Db        233 NGSHPSSSEGIVSRTACAH-----------------------------------WSD 254

Qy        310 TCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALYPGTNYT 369
              | ||                                          : |: :   ||  |
Db        255 HCCRW-----------------------------------------STEIQVKACPGGFYI 274

Qy        370 VNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRSR 418
              |: | ||  ::       |    :|:      :| |:  | | |
Db        275 YNL-TEPPECNLAYC-------TDPSSVEGTCEECRVDEDCISDNGRWR 315


RESULT 13
A35626
```

```
transforming growth factor beta-1-binding protein - human
C;Species: Homo sapiens (man)
C;Date: 21-Sep-1990 #sequence_revision 21-Sep-1990 #text_change 11-Jan-2000
C;Accession: A35626
R;Kanzaki, T.; Olofsson, A.; Moren, A.; Wernstedt, C.; Hellman, U.; Miyazono, K.; Claess
Cell 61, 1051-1061, 1990
A;Title: TGF-beta1 binding protein: a component of the large latent complex of TGF-beta1
A;Reference number: A35626; MUID:90275601; PMID:2350783
A;Accession: A35626
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1394 <KAN>
A;Cross-references: GB:M34057; NID:g339547; PIDN:AAA61160.1; PID:g339548
C;Superfamily: unassigned EGF-related proteins; EGF homology
C;Keywords: alternative splicing
F;750-791/Domain: EGF homology <EGF>

  Query Match              6.5%;  Score 260.5;  DB 2;  Length 1394;
  Best Local Similarity   27.6%;  Pred. No. 2e-08;
  Matches   82;  Conservative   32;  Mismatches  102;  Indels   81;  Gaps   16;

Qy         34 DGLDVCATCH--EHATCQQREGKKICICNYGFVGNG-RTQCVDKNECQFGATLVCGNHTS 90
              : :| |    |   |  |:  ||   |:|: |:  :|    | | |||    : ||
Db        792 EDIDECQHRHLCAHGQCRNTEGSFQCVCDQGYRASGLGDHCEDINECLEDKS-VC-QRGD 849

Qy         91 CHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFEC 150
              | || | : | | :|::  ::|||        | ||:|||  |||   | |:|| ||| |
Db        850 CINTAGSYDCTCPDGFQ-LDDNKT--------CQDINECEHPGLCGPQGECLNTEGSFHC 900

Qy        151 YCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACREGF 210
              |  |         |  : |  :| :||      | | :    :  :      |  | :|
Db        901 VCQQG---------FSISADGRTCEDIDECVNNTVCDSH---GFCDNTAGSFRCLCYQG- 947

Qy        211 FSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGF 270
              |  |:|      | |        | ::               |    | |    || | |
Db        948 FQAPQD------GQG--------CVDV---------------NECELLSG----VCGEAF 974

Qy        271 -ESPGGKITSVCTEK--------GTWRESTLCTEI-----------LTKINDVSL 306
               |:  |    || ::        |  |  | |  ::              :|| ||
Db        975 CENVEGSFLCVCADENQEYSPMTGQCRSRTSTDLDVDVDQPKEEKKECYYNLNDASL 1031


RESULT 14
A38261
masking protein precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 21-Jan-2000
C;Accession: A38261
R;Tsuji, T.; Okada, F.; Yamaguchi, K.; Nakamura, T.
Proc. Natl. Acad. Sci. U.S.A. 87, 8835-8839, 1990
A;Title: Molecular cloning of the large subunit of transforming growth factor type beta
A;Reference number: A38261; MUID:91062373; PMID:2247454
A;Accession: A38261
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1712 <TSU>
A;Cross-references: GB:M55431; NID:g207285; PIDN:AAA42235.1; PID:g207286
C;Superfamily: unassigned EGF-related proteins; EGF homology; LDL receptor ligand-bindin
F;911-947/Domain: EGF homology <EGF>

  Query Match              6.2%;  Score 251;  DB 2;  Length 1712;
  Best Local Similarity   26.4%;  Pred. No. 1e-07;
  Matches   79;  Conservative   30;  Mismatches  104;  Indels   86;  Gaps   14;

Qy         34 DGLDVCATCH--EHATCQQREGKKICICNYGF----VGNGRTQCVDKNECQFGATLVCGN 87
              : :| |    |   |  |:  ||   |:|| |:    :|:    | | |||   ::: |
Db       1111 EDIDECEHRHLCSHGQCRNTEGSFQCLCNQGYRASVLGD---HCEDINECLEDSSVCQGG 1167

Qy         88 HTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGS 147
                 | || | : | | :| :          ||   | ||:||   |||   | |:|| ||
Db       1168 --DCINTAGSYDCTCPDGLQL---------NDNKGCQDINECAQPGLCAPHGECLNTQGS 1216

Qy        148 FECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACR 207
```

```
Db       1217 FHCVCEQG---------FSISADGRTCEDIDECVNNTVCDSH---GFCDNTAGSFRCLCY 1264

Qy        208 EGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQ 267
              :| |  |:|      | |        | ::               |    | |    ||
Db       1265 QG-FQAPQD------GQG--------CVDV---------------NECELLSG----VCG 1290

Qy        268 EGF-ESPGGKITSVCTEK-------------------GTWRESTLTCTEILTKINDVSL 306
              | | |:  |    || ::                   |  |:     |    :|| ||
Db       1291 EAFCENVEGSFLCVCADENQEYSPMTGQCRSRATEDSGVDRQPKEEKKECYYNLNDASL 1349


RESULT 15
MMHUND
nidogen precursor - human
N;Alternate names: entactin
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Oct-2000
C;Accession: A33322; A32437; A61367
R;Nagayoshi, T.; Sanborn, D.; Hickok, N.J.; Olsen, D.R.; Fazio, M.J.; Chu, M.L.; Knowltc
DNA 8, 581-594, 1989
A;Title: Human nidogen: complete amino acid sequence and structural domains deduced from
A;Reference number: A33322; MUID:90091745; PMID:2574658
A;Accession: A33322
A;Molecule type: mRNA
A;Residues: 1-1247 <NAG>
A;Cross-references: EMBL:M30269
R;Olsen, D.R.; Nagayoshi, T.; Fazio, M.; Mattei, M.G.; Passage, E.; Weil, D.; Timpl, R.;
Am. J. Hum. Genet. 44, 876-885, 1989
A;Title: Human nidogen: cDNA cloning, cellular expression, and mapping of the gene to ch
A;Reference number: A32437; MUID:89270475; PMID:2471408
A;Accession: A32437
A;Molecule type: mRNA
A;Residues: 667-1247 <OLS>
A;Cross-references: EMBL:M27445; NID:g602466; PIDN:AAA57261.1; PID:g602467
A;Note: the authors translated the codon AAG for residue 966 as Cys
R;Fazio, M.J.; O'Leary, J.; Kaehaeri, V.M.; Chen, Y.Q.; Saitta, B.; Uitto, J.
J. Invest. Dermatol. 97, 281-285, 1991
A;Title: Human nidogen gene: structural and functional characterization of the 5'-flanki
A;Reference number: A61367; MUID:91302882; PMID:1906509
A;Accession: A61367
A;Molecule type: DNA
A;Residues: 1-28 <FAZ>
C;Comment: This protein is a basement membrane glycoprotein that forms a complex with la
C;Genetics:
A;Gene: GDB:NID
A;Cross-references: GDB:120236; OMIM:131390
A;Map position: 1q43-1q43
C;Superfamily: nidogen; EGF homology; LDL receptor YWTD-containing repeat homology; thyr
C;Keywords: basement membrane; beta-hydroxyasparagine; calcium binding; cell binding; co
protein
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1247/Product: nidogen #status predicted <MAT>
F;390-425/Domain: EGF homology <EG1>
F;672-708/Domain: EGF homology <EG2>
F;702-704/Region: cell attachment (R-G-D) motif
F;714-750/Domain: EGF homology <EG3>
F;762-800/Domain: EGF homology <EG4>
F;806-839/Domain: EGF homology <EG5>
F;849-919/Domain: thyroglobulin type I repeat homology <THY1>
F;990-1032/Domain: LDL receptor YWTD-containing repeat homology <YW1>
F;1033-1075/Domain: LDL receptor YWTD-containing repeat homology <YW2>
F;1076-1120/Domain: LDL receptor YWTD-containing repeat homology <YW3>
F;1121-1160/Domain: LDL receptor YWTD-containing repeat homology <YW4>
F;1161-1197/Domain: LDL receptor YWTD-containing repeat homology <YW5>
F;1212-1243/Domain: EGF homology <EG6>
F;289,296/Binding site: sulfate (Tyr) (covalent) #status predicted
F;729,819/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted
F;756/Cross-link: isopeptide (Gln) (interchain to Lys N6-amino of laminin) #status predi
F;1137/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match              6.1%;  Score 246.5;  DB 1;  Length 1247;
  Best Local Similarity   22.1%;  Pred. No. 1.3e-07;
```

Matches 138; Conservative 56; Mismatches 192; Indels 239; Gaps 30;

```
Qy      32 GPDGLDVCATCHEHATCQQREGKK---ICICNYGFVGNGRTQCVDKNECQFGATLVCGNH 88
           |  | |  | |        | | :    | |: || |:||| | | :||    : |||:|
Db     675 GTHGCDTNAAC--------RPGPRTQFTCECSIGFRGDGRT-CYDIDECSEQPS-VCGSH 724

Qy      89 TSCHNTPGGFYCICLEGY-------------RATNNNKTFIPN---------------- 118
           | |:| || | | |:|||              |  |  :| : |
Db     725 TICNNHPGTFRCECVEGYQFSDEGTCVAVVDQRPINYCETGLHNCDIPQRAQCIYTGGSS 784

Qy     119 -----------DGTFCTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGY----------- 156
                      ||  | |:|||: |  |     | || ||| | |  ||
Db     785 YTCSCLPGFSGDGQACQDVDECQPS-RCHPDAFCYNTPGSFTCQCKPGYQGDGFRCVPGE 843

Qy     157 ------------------------LPRNGPEPFHPTTDA------TSC----------- 174
                                   :|   |  | |  ||      | |
Db     844 VEKTRCQHEREHILGAAGATDPQRPIP---PGLFVPECDAHGHYAPTQCHGSTGYCWCVD 900

Qy     175 --------TEIDCG-TPP-------------------EVPDG-YIIGNYTSSL------ 198
                   |    | |||                   :| | :::   |  :
Db     901 RDGREVEGTRTRPGMTPPCLSTVAPPIHQGPAVPTAVIPLPPGTHLLFAQTGKIERLPLE 960

Qy     199 GSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGN---------PPEMRHAI 249
           |: :|    : |  ||   :   ||              :| :          | :  |
Db     961 GNTMRKTEAKAFLHVPAKVI---IGLA-----------FDCVDKMVYWTDITEPSIGRAS 1006

Qy     250 LVGNHSSRLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFND 309
           |   |    ||    : ::   || |          | :| |   |       |:  :
Db    1007 L---H----GGEPTTIIRQDLGSPEGIAVDWLGRNIFWTDSNLDRIE-------VAKLDG 1052

Qy     310 TCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNL-TTD--SRTPEVCLALYPGT 366
           |  |    :  :||:     |::|              ||  ||     |::  :   ||
Db    1053 TQRRVLFETDLVNPRGIVTDSVRG---------------NLYWTDWNRDNPKIETSYMDGT 1098

Qy     367 NYTVNISTAPPRRSMPAVIGFQTAEVDLLEWDGSFNISIFNETCLKLNRRSRKVGSEHM- 425
           |  : :          :|  : |       |   |    |        ||  :: ||:   | :
Db    1099 NRRILVQ---DDLGLPNGLTFDAFSSQLCWVDAGTN----RAECLNPSQPSRRKALEGLQ 1151

Qy     426 YQFTVLGQRWYLANFSHATSFNFTTREQVPVCLDL-----------YPTTDYTVNVTL 473
           | | |          |:  :  ||  :   || |||             |  | :   |
Db    1152 YPFAVT---------SYGKNLYFTDWKMNSVVALDLAISKETDAFQPHKQTRLYGITTAL 1202

Qy     474 LRSPKRHS---------VQITIATP 489
            : |: |:           : :|||
Db    1203 SQCPQGHNYCSVNNGGCTHLCLATP 1227
```

Search completed: February 6, 2004, 17:39:37
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 6, 2004, 17:36:49 ; Search time 22 Seconds
(without alignments)
1436.646 Million cell updates/sec

Title: US-10-066-269-58
Perfect score: 4029
Sequence: 1 MGRGPWDAGPSRRLLPLLLL.....VGLGSLAVVIILTFLSFSAV 747

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	284	7.0	835	4	US-09-284-819-6	Sequence 6, Appli
2	284	7.0	835	4	US-09-262-537-12	Sequence 12, Appl
3	274.5	6.8	886	3	US-09-110-116-3	Sequence 3, Appli
4	274.5	6.8	1833	3	US-08-479-722B-2	Sequence 2, Appli
5	274.5	6.8	1833	5	PCT-US95-02251-18	Sequence 18, Appl
6	260.5	6.5	1394	6	5177197-30	Patent No. 5177197
7	236	5.9	387	2	US-08-884-072-5	Sequence 5, Appli
8	236	5.9	387	2	US-08-833-963C-9	Sequence 9, Appli
9	236	5.9	387	3	US-08-980-514-3	Sequence 3, Appli
10	236	5.9	387	4	US-09-212-168-5	Sequence 5, Appli
11	236	5.9	1251	5	PCT-US95-02251-3	Sequence 3, Appli
12	236	5.9	1252	1	US-08-199-780-3	Sequence 3, Appli
13	236	5.9	1252	2	US-08-316-650-3	Sequence 3, Appli
14	236	5.9	1253	3	US-08-479-722B-4	Sequence 4, Appli
15	234	5.8	2199	4	US-08-793-273C-2	Sequence 2, Appli
16	234	5.8	2199	5	PCT-US95-11684-2	Sequence 2, Appli
17	231.5	5.7	2321	4	US-09-230-652-2	Sequence 2, Appli
18	231	5.7	1964	4	US-09-467-997-1	Sequence 1, Appli
19	230	5.7	2523	1	US-08-185-432-18	Sequence 18, Appl
20	230	5.7	2523	4	US-08-899-232-3	Sequence 3, Appli
21	228.5	5.7	816	2	US-08-820-170A-37	Sequence 37, Appl
22	228.5	5.7	816	3	US-09-055-699-37	Sequence 37, Appl
23	228.5	5.7	816	3	US-09-273-565-37	Sequence 37, Appl
24	228.5	5.7	816	4	US-09-565-538-37	Sequence 37, Appl
25	228.5	5.7	816	4	US-09-661-468-37	Sequence 37, Appl
26	228.5	5.7	816	4	US-09-976-165-37	Sequence 37, Appl
27	222	5.5	2703	1	US-08-185-432-19	Sequence 19, Appl
28	222	5.5	2703	4	US-08-899-232-4	Sequence 4, Appli
29	219	5.4	2556	1	US-08-185-432-17	Sequence 17, Appl
30	219	5.4	2556	4	US-08-899-232-2	Sequence 2, Appli
31	218	5.4	652	6	5258288-4	Patent No. 5258288
32	217.5	5.4	1810	4	US-08-793-273C-4	Sequence 4, Appli
33	217.5	5.4	1810	5	PCT-US95-11684-4	Sequence 4, Appli
34	216	5.4	2556	1	US-08-083-590A-20	Sequence 20, Appl
35	216	5.4	2556	3	US-08-532-384-20	Sequence 20, Appl
36	215	5.3	635	1	US-07-907-190-1	Sequence 1, Appli
37	215	5.3	635	1	US-07-985-691-2	Sequence 2, Appli
38	215	5.3	635	1	US-08-436-804-2	Sequence 2, Appli
39	215	5.3	635	1	US-08-267-387-2	Sequence 2, Appli
40	215	5.3	676	1	US-08-282-141-4	Sequence 4, Appli
41	215	5.3	676	1	US-08-435-434-3	Sequence 3, Appli
42	215	5.3	676	1	US-08-435-436-3	Sequence 3, Appli
43	215	5.3	676	2	US-08-438-863-3	Sequence 3, Appli
44	215	5.3	676	2	US-08-438-864-3	Sequence 3, Appli
45	215	5.3	676	3	US-08-438-862-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-284-819-6
; Sequence 6, Application US/09284819
; Patent No. 6365712
; GENERAL INFORMATION:
; APPLICANT: Kelly, Kathleen
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by The Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: Methods and Compositions for Inhibiting Inflammation
; TITLE OF INVENTION: and Angiogenesis Comprising a Mammalian CD97 Alpha
; TITLE OF INVENTION: Subunit
; FILE REFERENCE: 015280-263100US
; CURRENT APPLICATION NUMBER: US/09/284,819
; CURRENT FILING DATE: 1999-08-20
; EARLIER APPLICATION NUMBER: US 60/027,871
; EARLIER FILING DATE: 1996-10-25
; EARLIER APPLICATION NUMBER: WO PCT/US97/19772
; EARLIER FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 835
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CD97amino acid sequence encoded by full-length
; OTHER INFORMATION: clone pAT276
US-09-284-819-6

Query Match 7.0%; Score 284; DB 4; Length 835;
Best Local Similarity 21.8%; Pred. No. 2.2e-15;
Matches 168; Conservative 88; Mismatches 243; Indels 270; Gaps 40;

```
Qy          29 GAPGPDGLDVCATCHEHATCQQREGKKICICNYGF------VGNGRTQCVDKNECQFGAT 82
               ||    |       | ::::|       | || ||      :      | | |||   :
Db          17 GAETQDSRGCARWCPQNSSCVNATA---CRCNPGFSSFSEIITTPTETCDDINECATPSK 73

Qy          83 LVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVS-GLCRHGGRC 141
               : ||  : | || | : |:|  ||   :  |||        | |:|||: :  ||:  | |
Db          74 VSCGKFSDCWNTEGSYDCVCSPGYEPVSGTKTFKNESENTCQDVDECQQNPRLCKSYGTC 133

Qy         142 VNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGS- 200
               ||| ||: | |: |:  :  ||    ||   ||   |  |      : : |    :||
Db         134 VNTLGSYTCQCLPGF--KFIPEDPKVCTDVNECTS---GQNPCHSSTHCLNN----VGSY 184

Qy         201 QVRYACREGFFSV------PEDTVSSCTGLGTWESPKLHCQEIN-CGNPPEMRHAILVGN 253
               | |  || |:  :      | :||  |  :    | :  |     | |          ||:
Db         185 QCR--CRPGWQPIPGSPNGPNNTV--CEDVDECSSGQHQCDSSTVCFN--------TVGS 232
```

```
Qy        254 HSSRLGGVARYVCQEGFES----PGGKITSVCTEK--GTWRE----STLTCTEILTKIND 303
Db        233 YSCR--------CRPGWKPRHGIPNNQKDTVCEDMTFSTWTPPPGVHSQTLSRFFDKVQD 284

Qy        304 VSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALY 363
Db        285 LGRDSKT------SSAEVT--IQNVIKLVDELMEAPGDVE---------------------- 316

Qy        364 PGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRSRKVGSE 423
Db        317 ---------ALAPPVRHLIA-----TQLLSNLED-----------IMRILAKSLPKGP- 349

Qy        424 HMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTV--------NVTLLR 475
Db        350 ----------------FTYISPSN----------TELTLMIQERGDKNVTMGQ 376

Qy        476 SPKRHSVQITIAT-----PPAVK--QTISNISG-FNETCLRWRSIKTADMEEMYLFHIWG 527
Db        377 SSARMKLNWAVAAGAEDPGPAVAGILSIQNMTTLLANASLNLHSKKQAELEEIY------ 430

Qy        528 QRWYQKEFAQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQITEPP 587
Db        431 ----------------ESSIR----------GVQLRRLSAVNSIFLS--------- 451

Qy        588 LPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNAS 647
Db        452 ---------H---------NNTKELNSPI----------LFA---------FSHLE 470

Qy        648 DADGYVAAELLAKDV-PDDAMEI------PIGDRLYYGEYYNAP------LKRGSDYCII 694
Db        471 SSDGEAGRDPPAKDVMPGPRQELLCAFWKSDSDR---GGHWATEGCQVLGSKNGSTTCQC 527

Qy        695 LRITSEWNKVRRHSCAVWAQVKDSSLMLLQMAGVGLGSLAVVI-ILTFL 742
Db        528 SHLSS-FAILMAH-----YDVEDWKLTLITRVGLALSLFCLLLCILTFL 570
```

RESULT 2
US-09-262-537-12
; Sequence 12, Application US/09262537
; Patent No. 6479256
; GENERAL INFORMATION:
; APPLICANT: Hayflick, Joel
; TITLE OF INVENTION: Lectomedin Materials and Methods
; FILE REFERENCE: 27866/35307
; CURRENT APPLICATION NUMBER: US/09/262,537
; CURRENT FILING DATE: 1999-03-04
; EARLIER APPLICATION NUMBER: 60/076,782
; EARLIER FILING DATE: 1998-03-04
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 835
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-262-537-12

Query Match 7.0%; Score 284; DB 4; Length 835;
Best Local Similarity 21.8%; Pred. No. 2.2e-15;
Matches 168; Conservative 88; Mismatches 243; Indels 270; Gaps 40;

```
Qy         29 GAPGPDGLDVCATCHEHATCQQREGKKICICNYGF------VGNGRTQCVDKNECQFGAT 82
Db         17 GAETQDSRGCARWCPQNSSCVNATA---CRCNPGFSSFSEIITTPTETCDDINECATPSK 73

Qy         83 LVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVS-GLCRHGGRC 141
Db         74 VSCGKFSDCWNTEGSYDCVCSPGYEPVSGTKTFKNESENTCQDVDECQQNPRLCKSYGTC 133

Qy        142 VNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGS- 200
Db        134 VNTLGSYTCQCLPGF--KFIPEDPKVCTDVNECTS---GQNPCHSSTHCLNN----VGSY 184
```

```
Qy        201 QVRYACREGFFSV------PEDTVSSCTGLGTWESPKLHCQEIN-CGNPPEMRHAILVGN 253
Db        185 QCR--CRPGWQPIPGSPNGPNNTV--CEDVDECSSGQHQCDSSTVCFN--------TVGS 232

Qy        254 HSSRLGGVARYVCQEGFES----PGGKITSVCTEK--GTWRE----STLTCTEILTKIND 303
Db        233 YSCR--------CRPGWKPRHGIPNNQKDTVCEDMTFSTWTPPPGVHSQTLSRFFDKVQD 284

Qy        304 VSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALY 363
Db        285 LGRDSKT------SSAEVT--IQNVIKLVDELMEAPGDVE---------------------- 316

Qy        364 PGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRSRKVGSE 423
Db        317 ---------ALAPPVRHLIA-----TQLLSNLED-----------IMRILAKSLPKGP- 349

Qy        424 HMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTV--------NVTLLR 475
Db        350 ----------------FTYISPSN----------TELTLMIQERGDKNVTMGQ 376

Qy        476 SPKRHSVQITIAT-----PPAVK--QTISNISG-FNETCLRWRSIKTADMEEMYLFHIWG 527
Db        377 SSARMKLNWAVAAGAEDPGPAVAGILSIQNMTTLLANASLNLHSKKQAELEEIY------ 430

Qy        528 QRWYQKEFAQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISLTTQITEPP 587
Db        431 ----------------ESSIR----------GVQLRRLSAVNSIFLS--------- 451

Qy        588 LPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLVLPLALQSTFSCDSEGASSFFSNAS 647
Db        452 ---------H---------NNTKELNSPI----------LFA---------FSHLE 470

Qy        648 DADGYVAAELLAKDV-PDDAMEI------PIGDRLYYGEYYNAP------LKRGSDYCII 694
Db        471 SSDGEAGRDPPAKDVMPGPRQELLCAFWKSDSDR---GGHWATEGCQVLGSKNGSTTCQC 527

Qy        695 LRITSEWNKVRRHSCAVWAQVKDSSLMLLQMAGVGLGSLAVVI-ILTFL 742
Db        528 SHLSS-FAILMAH-----YDVEDWKLTLITRVGLALSLFCLLLCILTFL 570
```

RESULT 3
US-09-110-116-3
; Sequence 3, Application US/09110116
; Patent No. 6013479
; GENERAL INFORMATION:
; APPLICANT: Xu, Hong
; APPLICANT: Cohan, Victoria L.
; APPLICANT: Stuart, Susan G.
; TITLE OF INVENTION: HUMAN EMR1-LIKE G PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTOR
; FILE REFERENCE: PF-0550 US
; CURRENT APPLICATION NUMBER: US/09/110,116
; CURRENT FILING DATE: 1998-07-02
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 886
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
; FEATURE:
; OTHER INFORMATION: 784994, GenBank
US-09-110-116-3

Query Match 6.8%; Score 274.5; DB 3; Length 886;
Best Local Similarity 22.8%; Pred. No. 1.5e-14;
Matches 89; Conservative 53; Mismatches 125; Indels 123; Gaps 17;

```
Qy         42 CHEHATCQQREGKKICICNYGFV-GNGR-------TQCVDKNECQFGATLVCGNHTSCHN 93
Db         41 CPAYATCTNTVDSYYCTCKQGFLSSNGQNHFKDPGVRCKDIDECS-QSPQPCGPNSSCKN 99

Qy         94 TPGGFYCICLEGYRATNNNKTFIPNDGTF-CTDIDECEVSGLCRHGGRCVNTHGSFECYC 152
```

```
                 | : | ||:|: :   |       | | ||||:||  | :|     |||: ||: | |
Db        100 LSGRYKCSCLDGFSSPTGNDWVPGKPGNFSCTDINECLTSRVCPEHSDCVNSMGSYSCSC 159

Qy        153 MDGYLPRNGPEPFHPTTDATSCTEI-DCGTPPEVPD----GYIIGNYTSSLGSQVRYACR 207
               |:: ||           ::| :: :|  |   |:       :|||:          |
Db        160 QVGFISRN-----------STCEDVNECADPRACPEHATCNNTVGNYSC--------FCN 200

Qy        208 EGFFSVPEDTVSSCTGLGTWESPKLHCQEIN-----------CGNPPEMRHAILVGNHSS 256
               || |       || ||     |  |::|:           | | |
Db        201 PGFES--SSGHLSCQGL------KASCEDIDECTEMCPINSTCTNTP------------ 239

Qy        257 RLGGVARYVCQEGFESPGGKI----------------------TSVCTE--------- 283
                 |   |  ||    |::                      |:||
Db        240 ---GSYFCTCHPGFAPSSGQLNFTDQGVECRDIDECRQDPSTCGPNSICTNALGSYSCGC 296

Qy        284 --------KGTWRESTLTCTEILTKINDVSLFNDTCVRWQINSRRINP-------KISYV 328
                    :|: ::   :|  :| |  :  : ::  ::       : |        :|: :
Db        297 IVGFHPNPEGSQKDGNFSCQRVLFKCKEDVIPDNKQIQQCQEGTAVKPAYVSFCAQINNI 356

Qy        329 ISIKGQRLDPMESVREETVNL--TTDSRTP 356
               |:    || :   :   |:|  ||:|  |
Db        357 FSV----LDKVCENKTTVVSLKNTTESFVP 382
```

RESULT 4
US-08-479-722B-2
; Sequence 2, Application US/08479722B
; Patent No. 6074840
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Yin, Wushan
; TITLE OF INVENTION: LATENT TGF(BINDING PROTEIN (LTBP)
; TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Williams, Morgan & Amerson
; STREET: 7676 Hillmont, Suite 250
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,722B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US95/02251
; FILING DATE: 21-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fussey, Shelley P.M.
; REGISTRATION NUMBER: 39,458
; REFERENCE/DOCKET NUMBER: 4100.000500/FUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 934-7000
; TELEFAX: (713) 934-7011
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1833 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-479-722B-2

```
Query Match             6.8%;  Score 274.5;  DB 3;  Length 1833;
Best Local Similarity  26.4%;  Pred. No. 4.5e-14;
Matches   81;  Conservative   28;  Mismatches   97;  Indels  101;  Gaps   16;

Qy         31 PGPDG-----LDVC----ATCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGA 81
              | | |     :| |     ::|     |:  ||   |:|: ||     | | | |||
Db       1137 PNPLGNRCEDVDECEGPQSSC-RGGECKNTEGSYQCLCHQGFQLVNGTMCEDVNECV--G 1193

Qy         82 TLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRC 141
                 |  | | |: | |:|:|  |: :           || | |:|||  :  |  ||| |
Db       1194 EEHCAPHGECLNSLGSFFCLCAPGFASAEG--------GTRCQDVDECAATDPC-PGGHC 1244

Qy        142 VNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEID-------------CGTPPEVPD 187
              ||| ||| | |                | |: |:  | :||                |  |
Db       1245 VNTEGSFSCLCETA--------SFQPSPDSGECLDIDECEDREDPVCGAWRCENSP---- 1292

Qy        188 GYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRH 247
                |:|  |      |: ||:  |                                :| :  |  :
Db       1293 ----GSYRCILD------CQPGFYVAPNG--------------------DCIDIDECAN 1321

Qy        248 AILVGNHS--SRLGGVARYVCQEGFE-SPGGKITSVCTEKGTWRESTLTCTEILTKINDV 304
               : |||        |  | :| :||| || |           |     | :    :|:
Db       1322 DTVCGNHGFCDNTDGSFRCLCDQGFETSPSG----------WE-----CVD----VNEC 1361

Qy        305 SLFNDTC 311
              |   |
Db       1362 ELMMAVC 1368
```

RESULT 5
PCT-US95-02251-18
; Sequence 18, Application PC/TUS9502251
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
; TITLE OF INVENTION: CELLS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02251
; FILING DATE: CONCURRENTLY HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 18-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC009P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1833 amino acids

```
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-02251-18

  Query Match              6.8%;  Score 274.5;  DB 5;  Length 1833;
  Best Local Similarity   26.4%;  Pred. No. 4.5e-14;
  Matches   81;  Conservative   28;  Mismatches   97;  Indels  101;  Gaps   16;

Qy         31 PGPDG-----LDVC----ATCHEHATCQQREGKKICICNYGFVGNGRTQCVDKNECQFGA 81
              | | |     :| |    ::|      |:  ||    |:|: ||     | | | |||
Db       1137 PNPLGNRCEDVDECEGPQSSC-RGGECKNTEGSYQCLCHQGFQLVNGTMCEDVNECV--G 1193

Qy         82 TLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRC 141
              |  |  | |: | | |:|:|  |: :           || | |:|||  :  |  || |
Db       1194 EEHCAPHGECLNSLGSFFCLCAPGFASAEG--------GTRCQDVDECAATDPC-PGGHC 1244

Qy        142 VNTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEID--------------CGTPPEVPD 187
              ||| ||| | |              | |: |:  | :||              |   |
Db       1245 VNTEGSFSCLCETA--------SFQPSPDSGECLDIDECEDREDPVCGAWRCENSP---- 1292

Qy        188 GYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRH 247
              |:|   |      |: ||:  |                     :| :  |  :
Db       1293 ----GSYRCILD------CQPGFYVAPNG--------------------DCIDIDECAN 1321

Qy        248 AILVGNHS--SRLGGVARYVCQEGFE-SPGGKITSVCTEKGTWRESTLTCTEILTKINDV 304
               : |||       |  | :| :||| || |           |      | :    :|:
Db       1322 DTVCGNHGFCDNTDGSFRCLCDQGFETSPSG----------WE-----CVD----VNEC 1361

Qy        305 SLFNDTC 311
              |   |
Db       1362 ELMMAVC 1368


RESULT 6
5177197-30
;Patent No. 5177197
;    APPLICANT: KANZAKI, TETSUTO;OLOFSSON, ANDERS;MOREN, ANITA;
;WERNSTEDT, CHRISTER;HELLMAN, ULF;MIYAZONO, KOHEI;CLAESSON-WELSH,
;LENA;HELDIN, CARL-HENRIK
;    TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
;HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
;    NUMBER OF SEQUENCES: 53
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/487,343
;      FILING DATE: 27-FEB-1990
;SEQ ID NO:30:
;      LENGTH: 1394
5177197-30

  Query Match              6.5%;  Score 260.5;  DB 6;  Length 1394;
  Best Local Similarity   27.6%;  Pred. No. 4.7e-13;
  Matches   82;  Conservative   32;  Mismatches  102;  Indels   81;  Gaps   16;

Qy         34 DGLDVCATCH--EHATCQQREGKKICICNYGFVGNG-RTQCVDKNECQFGATLVCGNHTS 90
              : :|  |    |  |  |:  ||   |:|: |:  :|   | | |||    : ||
Db        792 EDIDECQHRHLCAHGQCRNTEGSFQCVCDQGYRASGLGDHCEDINECLEDKS-VC-QRGD 849

Qy         91 CHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFEC 150
              | || | : | | :|::  ::|||        | ||:|||  |||   | |:|| ||| |
Db        850 CINTAGSYDCTCPDGFQ-LDDNKT--------CQDINECEHPGLCGPQGECLNTEGSFHC 900

Qy        151 YCMDGYLPRNGPEPFHPTTDATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACREGF 210
              | |          | : |  :| :||      | | :    :  :      |  | :|
Db        901 VCQQG---------FSISADGRTCEDIDECVNNTVCDSH---GFCDNTAGSFRCLCYQG- 947

Qy        211 FSVPEDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGF 270
              |  |:|     | |       | ::              |    | |    || | |
Db        948 FQAPQD------GQG--------CVDV--------------NECELLSG----VCGEAF 974

Qy        271 -ESPGGKITSVCTEK--------GTWRESTLTCTEI------------LTKINDVSL 306
               /   |:  |    ||  ::       |  |  | |  ::              :|| ||
Db        975 CENVEGSFLCVCADENQEYSPMTGQCRSRTSTDLDVDVDQPKEEKKECYYNLNDASL 1031
```

```
RESULT 7
US-08-884-072-5
; Sequence 5, Application US/08884072
; Patent No. 5872234
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl J.
;    TITLE OF INVENTION:  HUMAN EXTRACELLULAR MATRIX PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/884,072
;      FILING DATE:  Herewith
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0333 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  458228
US-08-884-072-5

  Query Match              5.9%;  Score 236;  DB 2;  Length 387;
  Best Local Similarity   23.9%;  Pred. No. 8.4e-12;
  Matches   94;  Conservative   46;  Mismatches  139;  Indels  114;  Gaps   19;

Qy          3 RGPWD-----AGPSRRLLPLLLLLGLARGAAGAPGPD-----GLDVCA----TCHEHATC 48
              | | |     : || |:          : |||    :      :| |      |      |
Db         35 RNPADPQRIPSNPSHRI----------QCAAGYEQSEHNVCQDIDECTAGTHNCRADQVC 84

Qy         49 QQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRA 108
                |  | |  |:  |  |||| :||    |:     |  | |||| ||| |  |::
Db         85 INLRGSFACQCPPGYQKRGE-QCVDIDEC----TIPPYCHQRCVNTPGSFYCQCSPGFQL 139

Qy        109 TNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPT 168
              || |         | ||:||: |  |    :| |  ||| | |  ||           :
Db        140 AANNYT--------CVDINECDASNQCAQ--QCYNILGSFICQCNQGY---------ELS 180

Qy        169 TDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTW 227
              :|  :| :|| | |       |:      :    :  | :|:  |   |
Db        181 SDRLNCEDIDECRT-----SSYLCQYQCVNEPGKFSCMCPQGYQVVRSRT---------- 225

Qy        228 ESPKLHCQEIN-CGNPPEMRHAILVGNHSSRLGGVARYVCQEGF-ESPGGKITSVCTEKG 285
                  ||:|| |    | |   :  |:       |  ||: :  :|  :   ||
Db        226 ------CQDINECETTNECREDEMCWNYHGGFRCYPRNPCQDPYILTPENR--CVCP--- 274
```

```
Qy        286 TWRESTLTCTEI-------------------LTKINDVSLFNDTCVRWQINS------- 318
Db        275 ---VSNAMCRELPQSIVYKYMSIRSDRSVPSDIFQIQATTIYANTINTFRIKSGNENGEF 331

Qy        319 --RRINPKISYVISIKGQRLDPMESVREETVNL 349
Db        332 YLRQTSPVSAMLVLVKS-----LSGPREHIVDL 359
```

RESULT 8
US-08-833-963C-9
; Sequence 9, Application US/08833963C
; Patent No. 5916769
; GENERAL INFORMATION:
; APPLICANT: Olsen, et al.
; TITLE OF INVENTION: Extracellular/Epidermal Growth Factor
; TITLE OF INVENTION: HCABA58X
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,963C
; FILING DATE: 11-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US96/05033
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 387 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-833-963C-9

Query Match 5.9%; Score 236; DB 2; Length 387;
Best Local Similarity 23.9%; Pred. No. 8.4e-12;
Matches 94; Conservative 46; Mismatches 139; Indels 114; Gaps 19;

```
Qy          3 RGPWD-----AGPSRRLLPLLLLLGLARGAAGAPGPD-----GLDVCA----TCHEHATC 48
Db         35 RNPADPQRIPSNPSHRI----------QCAAGYEQSEHNVCQDIDECTAGTHNCRADQVC 84

Qy         49 QQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRA 108
Db         85 INLRGSFACQCPPGYQKRGE-QCVDIDEC----TIPPYCHQRCVNTPGSFYCQCSPGFQL 139

Qy        109 TNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPT 168
Db        140 AANNYT--------CVDINECDASNQCAQ--QCYNILGSFICQCNQGY---------ELS 180

Qy        169 TDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTW 227
Db        181 SDRLNCEDIDECRT-----SSYLCQYQCVNEPGKFSCMCPQGYQVVRSRT---------- 225
```

```
Qy        228 ESPKLHCQEIN-CGNPPEMRHAILVGNHSSRLGGVARYVCQEGF-ESPGGKITSVCTEKG 285
Db        226 ------CQDINECETTNECREDEMCWNYHGGFRCYPRNPCQDPYILTPENR--CVCP--- 274

Qy        286 TWRESTLTCTEI-------------------LTKINDVSLFNDTCVRWQINS------- 318
Db        275 ---VSNAMCRELPQSIVYKYMSIRSDRSVPSDIFQIQATTIYANTINTFRIKSGNENGEF 331

Qy        319 --RRINPKISYVISIKGQRLDPMESVREETVNL 349
Db        332 YLRQTSPVSAMLVLVKS-----LSGPREHIVDL 359
```

RESULT 9
US-08-980-514-3
; Sequence 3, Application US/08980514
; Patent No. 6004753
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN S1-5-ECMP-LIKE PROT
; TITLE OF INVENTION: EIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,514
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0436 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 387 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 458228
US-08-980-514-3

Query Match 5.9%; Score 236; DB 3; Length 387;
Best Local Similarity 23.9%; Pred. No. 8.4e-12;
Matches 94; Conservative 46; Mismatches 139; Indels 114; Gaps 19;

```
Qy          3 RGPWD-----AGPSRRLLPLLLLLGLARGAAGAPGPD-----GLDVCA----TCHEHATC 48
Db         35 RNPADPQRIPSNPSHRI----------QCAAGYEQSEHNVCQDIDECTAGTHNCRADQVC 84

Qy         49 QQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRA 108
Db         85 INLRGSFACQCPPGYQKRGE-QCVDIDEC----TIPPYCHQRCVNTPGSFYCQCSPGFQL 139
```

```
Qy        109 TNNNKTFIPNDGTFCTDIDECEVSGLCRHGCRCVNTHGSFECYCMDGYLPRNGPEPFHPT 168
              ||  |         | ||:||: |  |     :| |  ||| | |  ||          :
Db        140 AANNYT--------CVDINECDASNQCAQ--QCYNILGSFICQCNQGY---------ELS 180

Qy        169 TDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTW 227
              :|  :| :|| | |        |:      :   :     | :|:  |    |
Db        181 SDRLNCEDIDECRT-----SSYLCQYQCVNEPGKFSCMCPQGYQVVRSRT---------- 225

Qy        228 ESPKLHCQEIN-CGNPPEMRHAILVGNHSSNLGGVARYVCQEGF-ESPGGKITSVCTEKG 285
                ||:|| |    | |   :  |:         |  ||: :  :|  :   ||
Db        226 ------CQDINECETTNECREDEMCWNYHGGFRCYPRNPCQDPYILTPENR--CVCP--- 274

Qy        286 TWRESTLTCTEI-------------------LTKINDVSLFNDTCVRWQINS------- 318
                |   | |:                    : :|   ::: :|   ::| |
Db        275 ---VSNAMCRELPQSIVYKYMSIRSDRSVPSDIFQIQATTIYANTINTFRIKSGNENGEF 331

Qy        319 --RRINPKISYVISIKGQRLDPMESVREETVNL 349
                |: :|  : :: :|      :   ||  :|
Db        332 YLRQTSPVSAMLVLVKS-----LSGPREHIVDL 359


RESULT 10
US-09-212-168-5
; Sequence 5, Application US/09212168
; Patent No. 6303765
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl J.
;    TITLE OF INVENTION:  HUMAN EXTRACELLULAR MATRIX PROTEINS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/212,168
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/884,072
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0333 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  387 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  458228
US-09-212-168-5

  Query Match             5.9%;  Score 236;  DB 4;  Length 387;
  Best Local Similarity   23.9%;  Pred. No. 8.4e-12;
  Matches  94;  Conservative   46;  Mismatches  139;  Indels  114;  Gaps   19;
```

```
Qy          3 RGPWD-----AGPSRRLLPLLLLLGLARGAAGAPGPD-----GLDVCA----TCHEHATC 48
              | | |     : || |:          : |||    :     :| |      |      |
Db         35 RNPADPQRIPSNPSHRI----------QCAAGYEQSEHNVCQDIDECTAGTHNCRADQVC 84

Qy         49 QQREGKKICICNYGFVGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRA 108
                |  | |  |:   |  |||| :||    |:    |  | |||| ||| |   |::
Db         85 INLRGSFACQCPPGYQKRGE-QCVDIDEC----TIPPYCHQRCVNTPGSFYCQCSPGFQL 139

Qy        109 TNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCVNTHGSFECYCMDGYLPRNGPEPFHPT 168
                || |         | ||:||: |  |     :| |  ||| | |  ||          :
Db        140 AANNYT--------CVDINECDASNQCAQ--QCYNILGSFICQCNQGY---------ELS 180

Qy        169 TDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTW 227
              :|  :| :|| | |        |:      :   :     | :|:  |    |
Db        181 SDRLNCEDIDECRT-----SSYLCQYQCVNEPGKFSCMCPQGYQVVRSRT---------- 225

Qy        228 ESPKLHCQEIN-CGNPPEMRHAILVGNHSSRLGGVARYVCQEGF-ESPGGKITSVCTEKG 285
                ||:|| |    | |   :  |:         |  ||: :  :|  :   ||
Db        226 ------CQDINECETTNECREDEMCWNYHGGFRCYPRNPCQDPYILTPENR--CVCP--- 274

Qy        286 TWRESTLTCTEI-------------------LTKINDVSLFNDTCVRWQINS------- 318
                |   | |:                    : :|   ::: :|   ::| |
Db        275 ---VSNAMCRELPQSIVYKYMSIRSDRSVPSDIFQIQATTIYANTINTFRIKSGNENGEF 331

Qy        319 --RRINPKISYVISIKGQRLDPMESVREETVNL 349
                |: :|  : :: :|      :   ||  :|
Db        332 YLRQTSPVSAMLVLVKS-----LSGPREHIVDL 359


RESULT 11
PCT-US95-02251-3
; Sequence 3, Application PC/TUS9502251
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR STIMULATING BONE
;    TITLE OF INVENTION:  CELLS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States of America
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS/ASCII
;      SOFTWARE:  PatentIn Release #1.0, Version
;      SOFTWARE:  #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/02251
;      FILING DATE:  CONCURRENTLY HEREWITH
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/316,650
;      FILING DATE:  30-SEP-1994
;      CLASSIFICATION:
;      APPLICATION NUMBER:  US 08/199,780
;      FILING DATE:  18-FEB-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UMIC009P--
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1251 amino acids
;      TYPE:  amino acid
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-02251-3

  Query Match              5.9%;  Score 236;  DB 5;  Length 1251;
  Best Local Similarity   23.0%;  Pred. No. 4.8e-11;
  Matches  149;  Conservative   48;  Mismatches  218;  Indels  234;  Gaps   36;

Qy         31 PGPDGLDVCATCH---------EHATCQQ---------REGKKICI---------CNYGF 63
              |||       :||         :|  |              || ||:         || |:
Db        571 PGPSDY----SCHCNAGYRSHPQHRYCVDVNECEAEPCGPGKGICMNTGGSYNCHCNRGY 626

Qy         64 ---VGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDG 120
                || |   ||| |||     :||:   | | || : | |  |||      |   |
Db        627 RLHVGAGGRSCVDLNEC--AKPHLCGDGGFCINFPGHYKCNCYPGYRL----KASRP--- 677

Qy        121 TFCTDIDECEVSGLCRHGGRCVNTHGSFECY-CMDGYLPRNGPEPFHPTTDATSCTEIDC 179
                | |||||     |   |:| |  |||:|  |  ||  : |        |   |:|
Db        678 PICEDIDECRDPSTC-PDGKCENKPGSFKCIACQPGYRSQGG----GACRDVNECSE--- 729

Qy        180 GTP------PEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLH 233
              |||       ::|  |             | | :|  :       ||  :   |: |:
Db        730 GTPCSPGWCEKLPGSY-----------RCTCAQGIRT--RTGRLSCIDVDDCEAGKV- 773

Qy        234 CQEINCGNPP-EMRHAILVGNHSSR-----------------------LGGVARYVCQE 268
              ||:  | | |   :   | | | ||                       | | :|
Db        774 CQDGICTNTPGSFQCQCLSGYHLSRDRSRCEDIDECDFPAACIGGDCINTNGSYRCLCPL 833

Qy        269 GF---------------ESPG----------GKITSVCTEKGTWRESTLTCTEILT-- 299
              |               : ||          |   || |  |  :   | |:
Db        834 GHRLVGGRKCKKDIDECSQDPGLCLPHACENLQGSYVCVCDEGFTLTQDQHGCEEVEQPH 893

Qy        300 -KINDVSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEV 358
               |      |:||                 :||      | |:    |
Db        894 HKKECYLNFDDTVF-----------------------CDSV------LATNVTQQEC 921

Qy        359 CLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGS--FNISIFNETCLKLNR- 415
              | :|  |       |  |          : :||   |  ||    :      | |:  :|
Db        922 CCSLGAGWGDHCEIYPCPV---------YSSAEFHSLVPDGKRLHSGQQHCELCIPAHRD 972

Qy        416 -------------RSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLY 462
                          : | |:  |:       :|  |                 : |:|
Db        973 IDECILFGAEICKEGKCVNSQPGYECYCKQGFYYDGNL---------------LECVD-- 1015

Qy        463 PTTDYTVNVTLLRSPKRHSV-----QITIA-TPPAVKQTISNISGFNETCL---RW---- 509
                 |:  |  |  |: |     ::  | ||||       |     ||   ||
Db       1016 ------VDECLDESNCRNGVCENTWRLPCACTPPA------EYSPAQAQCLIPERWSTPQ 1063

Qy        510 RSIKTADMEEMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRP 558
              | :| |   |     :||  |       :||:     | | :    ||
Db       1064 RDVKCAGASEERTACVWGP-W----AGPALTFD-DCCCRQPRLGTQCRP 1106


RESULT 12
US-08-199-780-3
; Sequence 3, Application US/08199780
; Patent No. 5763416
;  GENERAL INFORMATION:
;    APPLICANT:  Bonadio, Jeffrey
;    APPLICANT:  Goldstein, Steven A.
;    TITLE OF INVENTION:  Gene Transfer Into Bone Cells
;    TITLE OF INVENTION:  And Tissues
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  TX
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/199,780
;      FILING DATE:  18-FEB-1994
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UMIC:002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 320-7200
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1252 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-199-780-3

  Query Match              5.9%;  Score 236;  DB 1;  Length 1252;
  Best Local Similarity   23.0%;  Pred. No. 4.8e-11;
  Matches  149;  Conservative   48;  Mismatches  218;  Indels  234;  Gaps   36;

Qy         31 PGPDGLDVCATCH---------EHATCQQ---------REGKKICI---------CNYGF 63
              |||       :||         :|  |              || ||:         || |:
Db        572 PGPSDY----SCHCNAGYRSHPQHRYCVDVNECEAEPCGPGKGICMNTGGSYNCHCNRGY 627

Qy         64 ---VGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDG 120
                || |   ||| |||     :||:   | | || : | |  |||      |   |
Db        628 RLHVGAGGRSCVDLNEC--AKPHLCGDGGFCINFPGHYKCNCYPGYRL----KASRP--- 678

Qy        121 TFCTDIDECEVSGLCRHGGRCVNTHGSFECY-CMDGYLPRNGPEPFHPTTDATSCTEIDC 179
                | |||||     |   |:| |  |||:|  |  ||  : |        |   |:|
Db        679 PICEDIDECRDPSTC-PDGKCENKPGSFKCIACQPGYRSQGG----GACRDVNECSE--- 730

Qy        180 GTP------PEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLH 233
              |||       ::|  |             | | :|  :       ||  :   |: |:
Db        731 GTPCSPGWCEKLPGSY-----------RCTCAQGIRT--RTGRLSCIDVDDCEAGKV- 774

Qy        234 CQEINCGNPP-EMRHAILVGNHSSR-----------------------LGGVARYVCQE 268
              ||:  | | |   :   | | | ||                       | | :|
Db        775 CQDGICTNTPGSFQCQCLSGYHLSRDRSRCEDIDECDFPAACIGGDCINTNGSYRCLCPL 834

Qy        269 GF---------------ESPG----------GKITSVCTEKGTWRESTLTCTEILT-- 299
              |               : ||          |   || |  |  :   | |:
Db        835 GHRLVGGRKCKKDIDECSQDPGLCLPHACENLQGSYVCVCDEGFTLTQDQHGCEEVEQPH 894

Qy        300 -KINDVSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEV 358
               |      |:||                 :||      | |:    |
Db        895 HKKECYLNFDDTVF-----------------------CDSV------LATNVTQQEC 922

Qy        359 CLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGS--FNISIFNETCLKLNR- 415
              | :|  |       |  |          : :||   |  ||    :      | |:  :|
Db        923 CCSLGAGWGDHCEIYPCPV---------YSSAEFHSLVPDGKRLHSGQQHCELCIPAHRD 973

Qy        416 -------------RSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLY 462
                          : | |:  |:       :|  |                 : |:|
Db        974 IDECILFGAEICKEGKCVNSQPGYECYCKQGFYYDGNL---------------LECVD-- 1016

Qy        463 PTTDYTVNVTLLRSPKRHSV-----QITIA-TPPAVKQTISNISGFNETCL---RW---- 509
                 |:  |  |  |: |     ::  | ||||       |     ||   ||
Db       1017 ------VDECLDESNCRNGVCENTWRLPCACTPPA------EYSPAQAQCLIPERWSTPQ 1064

Qy        510 RSIKTADMEEMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRP 558
              | :| |   |     :||  |       :||:     | | :    ||
Db       1065 RDVKCAGASEERTACVWGP-W----AGPALTFD-DCCCRQPRLGTQCRP 1107


RESULT 13
US-08-316-650-3
; Sequence 3, Application US/08316650
```

```
; Patent No. 5942496
;  GENERAL INFORMATION:
;    APPLICANT:  Bonadio, Jeffrey
;    APPLICANT:  Roessler, Blake J.
;    APPLICANT:  Goldstein, Steven A.
;    APPLICANT:  Lin, Wushan
;    TITLE OF INVENTION:  METHODS AND COMPOSITIONS
;    TITLE OF INVENTION:  FOR STIMULATING BONE CELLS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/316,650
;      FILING DATE:  30-SEP-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/199,780
;      FILING DATE:  30-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, David L.
;      REGISTRATION NUMBER:  32,165
;      REFERENCE/DOCKET NUMBER:  UMIC:008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (713) 789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1252 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-316-650-3

  Query Match             5.9%;  Score 236;  DB 2;  Length 1252;
  Best Local Similarity   23.0%;  Pred. No. 4.8e-11;
  Matches  149;  Conservative   48;  Mismatches  218;  Indels  234;  Gaps   36;

Qy         31 PGPDGLDVCATCH---------EHATCQQ---------REGKKICI---------CNYGF 63
              |||       :||         :|  |             || ||:         || |:
Db        572 PGPSDY----SCHCNAGYRSHPQHRYCVDVNECEAEPCGPGKGICMNTGGSYNCHCNRGY 627

Qy         64 ---VGNGRTQCVDKNECQFGATLVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDG 120
                || |   ||| |||      :||:   | | || : | |  |||     |   |
Db        628 RLHVGAGGRSCVDLNEC--AKPHLCGDGGFCINFPGHYKCNCYPGYRL----KASRP--- 678

Qy        121 TFCTDIDECEVSGLCRHGGRCVNTHGSFECY-CMDGYLPRNGPEPFHPTTDATSCTEIDC 179
               | |||||      |   |:| |  |||:|  |  ||  : |          |   |:|
Db        679 PICEDIDECRDPSTC-PDGKCENKPGSFKCIACQPGYRSQGG----GACRDVNECSE--- 730

Qy        180 GTP------PEVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTWESPKLH 233
              |||       ::|  |             |  | :|  :       ||  :   |: |:
Db        731 GTPCSPGWCEKLPGSY------------RCTCAQGIRT--RTGRLSCIDVDDCEAGKV- 774

Qy        234 CQEINCGNPP-EMRHAILVGNHSSR---------------------LGGVARYVCQE 268
              ||:  | | |   :   | | | ||                     |  | :|
Db        775 CQDGICTNTPGSFQCQCLSGYHLSRDRSRCEDIDECDFPAACIGGDCINTNGSYRCLCPL 834

Qy        269 GF---------------ESPG-----------GKITSVCTEKGTWRESTLTCTEILT-- 299
              |               : ||           |    || |  |  :    | |:
Db        835 GHRLVGGRKCKKDIDECSQDPGLCLPHACENLQGSYVCVCDEGFTLTQDQHGCEEVEQPH 894

Qy        300 -KINDVSLFNDTCVRWQINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEV 358
```

```
                |      |:||                              :||      | |:    |
Db        895 HKKECYLNFDDTVF-------------------------CDSV------LATNVTQQEC 922

Qy        359 CLALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGS--FNISIFNETCLKLNR- 415
              | :|  |      |   |          : :||   |  ||    :     | |:  :|
Db        923 CCSLGAGWGDHCEIYPCPV---------YSSAEFHSLVPDGKRLHSGQQHCELCIPAHRD 973

Qy        416 ------------RSRKVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLY 462
                         : | |:  |:      :|  |                  : |:|
Db        974 IDECILFGAEICKEGKCVNSQPGYECYCKQGFYYDGNL--------------LECVD-- 1016

Qy        463 PTTDYTVNVTLLRSPKRHSV-----QITIA-TPPAVKQTISNISGFNETCL---RW---- 509
                    |:  |  |  |: |     ::  | ||||       |     ||  .  ||
Db       1017 ------VDECLDESNCRNGVCENTWRLPCACTPPA------EYSPAQAQCLIPERWSTPQ 1064

Qy        510 RSIKTADMEEMYLFHIWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRP 558
              | :| |   |     :||  |      :||:     | | :    ||
Db       1065 RDVKCAGASEERTACVWGP-W----AGPALTFD-DCCCRQPRLGTQCRP 1107


RESULT 14
US-08-479-722B-4
; Sequence 4, Application US/08479722B
; Patent No. 6074840
;  GENERAL INFORMATION:
;    APPLICANT:  Bonadio, Jeffrey
;    APPLICANT:  Yin, Wushan
;    TITLE OF INVENTION:  LATENT TGF( BINDING PROTEIN (LTBP)
;    TITLE OF INVENTION:  GENES, COMPOSITIONS AND METHODS
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Williams, Morgan & Amerson
;      STREET:  7676 Hillmont, Suite 250
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77040
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/479,722B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US95/02251
;      FILING DATE:  21-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/316,650
;      FILING DATE:  30-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/199,780
;      FILING DATE:  18-FEB-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fussey, Shelley P.M.
;      REGISTRATION NUMBER:  39,458
;      REFERENCE/DOCKET NUMBER:  4100.000500/FUS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (713) 934-7000
;      TELEFAX:  (713) 934-7011
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1253 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-479-722B-4

  Query Match             5.9%;  Score 236;  DB 3;  Length 1253;
  Best Local Similarity   23.2%;  Pred. No. 4.8e-11;
  Matches  133;  Conservative   53;  Mismatches  206;  Indels  182;  Gaps   31;
```

```
Qy         27 AAGAPGPDGLDVC---ATCHEHATCQQREGKKICI-CNYGFVGNGRTQCVDKNECQFGAT 82
Db        675 ASRPPICEDIDECRDPSTCPD-GKCENKPGSFKCIACQPGYRSQGGGACRDVNECSEGTP 733

Qy         83 LVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPNDGTFCTDIDECEVSGLCRHGGRCV 142
Db        734 CSPG---WCENLPGSYRCTCAQGIRTRTGRLS--------CIDVDECEAGKVCQ-DGICT 781

Qy        143 NTHGSFECYCMDGYLPRNGPEPFHPTTDATSCTEID-CGTPPEVPDGYIIGNYTSSLGSQ 201
Db        782 NTPGSFQCQCLSGY---------HLSRDRSRCEDIDEC----DFPAACIGGDCINTNGS- 827

Qy        202 VRYACREGFFSVP----EDTVSSCTGLGTWESPKLHCQEINCGNPPEMRHAILVGNHSSR 257
Db        828 YRCLCPLGHRLVGGRKCKKDIDECS-----QDPGL-CLPHACEN--------LQGS---- 869

Qy        258 LGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEILT---KINDVSLFNDTCVRW 314
Db        870 ------YVC-------------VCDEGFTLTQDQHGCEEVEQPHHKKECYLNFDDTVF-- 908

Qy        315 QINSRRINPKISYVISIKGQRLDPMESVREETVNLTTDSRTPEVCLALYPGTNYTVNIST 374
Db        909 ------------------------CDSV------LATNVTQQECCCSLGAGWGDHCEIYP 938

Qy        375 APPRRSMPAVIGFQTAEVDLLEDDGS--FNISIFNETCLKLNR--------------RSR 418
Db        939 CPV---------YSSAEFHSLVPDGKRLHSGQQHCELCIPAHRDIDECILFGAEICKEGK 989

Qy        419 KVGSEHMYQFTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTVNVTLLRSPK 478
Db        990 CVNTQPGYECYCKQGFYYDGNL----------------LECVD--------VDECLDESNC 1026

Qy        479 RHSV-------QITIATPPAVKQTISNISGFNETCL---RW----RSIKTADMEEMYLFH 524
Db       1027 RNGVCENTRGGYRCACTPPA------EYSPAQAQCLIPERWSTPQRDVKCAGASEERTAC 1080

Qy        525 IWGQRWYQKEFAQEMTFNISSSSRDPEVCLDLRP 558
Db       1081 VWGP-W----AGPALTFD-DCCCRQPRLGTQCRP 1108


RESULT 15
US-08-793-273C-2
; Sequence 2, Application US/08793273C
; Patent No. 6482410
; GENERAL INFORMATION:
;  APPLICANT: Crossin, Kathryn L.
;  APPLICANT:  Phillips, Greg
;  APPLICANT:  Prieto, Anne L.
;  TITLE OF INVENTION: CYTOTACTIN DERIVATIVES THAT STIMULATE ATTACHMENT AND
;  TITLE OF INVENTION:  NEURITE OUTGROWTH, AND METHODS OF MAKING SAME
;  FILE REFERENCE: BEC0022S
;  CURRENT APPLICATION NUMBER: US/08/793,273C
;  CURRENT FILING DATE:  1997-05-22
;  PRIOR APPLICATION NUMBER: PCT/US95/11684
;  PRIOR FILING DATE: 1995-09-14
;  PRIOR APPLICATION NUMBER: 08/308,359
;  PRIOR FILING DATE: 1994-09-16
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 2199
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-08-793-273C-2

  Query Match             5.8%;  Score 234;  DB 4;  Length 2199;
  Best Local Similarity  19.8%;  Pred. No. 1.6e-10;
  Matches  152;  Conservative   74;  Mismatches  217;  Indels  326;  Gaps   33;

Qy         29 GAPGPD-GLDVCA-TCHEHATCQQREGKKICICNYGFVG-----------NGRTQCVD- 73
Db        334 GFTGEDCGKPTCPHACHTQGRCE--EGQ--CVCDEGFAGVDCSEKRCPADCHNRGRCVDG 389

Qy         74 KNECQFGAT------LVCGNHTSCHNTPGGFYCICLEGYRATNNNKTFIPND-------- 119
Db        390 RCECDDGFTGADCGELKCPNGCSGHGRCVNGQCVCDEGYTGEDCSQLRCPNDCHSRGRCV 449

Qy        120 -----------GTFCTDID-----------------------------------ECEVSGL 134
Db        450 EGKCVCEQGFKGYDCSDMSCPNDCHQHGRCVNGMCVCDDGYTGEDCRDRQCPRDCSNRGL 509

Qy        135 CRHG-------------------------GRCVNTHGSFECYCMDGYLPRNGPEPFHPTT 169
Db        510 CVDGQCVCEDGFTGPDCAELSCPNDCHGQGRCVNG----QCVCHEGFMGKDCKEQRCPS- 564

Qy        170 DATSCTEIDCGTPPEVPDGYIIGNYTSSLGSQVRYACREGF-------FSVPEDTVSSCT 222
Db        565 --------DCHGQGRCVDGQCI-------------CHEGFTGLDCGQHSCPSD----CN 598

Qy        223 GLGTWESPKLHCQEINCG------NPP-------------------EMR-HAILVGNHSS 256
Db        599 NLGQCVSGRCICNEGYSGEDCSEVSPPKDLVVTEVTEETVNLAWDNEMRVTEYLVVYTPT 658

Qy        257 RLGGVARYVCQEGFESPGGKITSVCTEKGTWRESTLTCTEILTKINDVSL---------- 306
Db        659 HEGGL-----EMQFRVPGDQTSTIIRELEPGVEYFIRVFAILENKKSIPVSARVATYLPA 713

Qy        307 --------FNDTCVR-----------WQINSRRIN-------------PKISY------ 327
Db        714 PEGLKFKSIKETSVEVEWDPLDIAFETWEIIFRNMNKEDEGEITKSLRRPETSYRQTGLA 773

Qy        328 --------------------VISIKGQRLDPMESVREETVNLTTDSRTPEVCLALYPGTN 367
Db        774 PGQEYEISLHIVKNNTRGPGLKRVTTTRLDAPSQIEVKDVTDTTALITWFKPLAEIDGIE 833

Qy        368 YTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNISIFNETCLKLNRRSRKVGSEHMYQ 427
Db        834 LTYGIKDVPGDRT----------TIDLTEDENQYSIG-----------NLKPDTE--YE 869

Qy        428 FTVLGQRWYLANFSHATSFNFTTREQVPVVCLDLYPTTDYTVNVTLLRSPKRHSVQITIA 487
Db        870 VSLISRRGDMS--SNPAKETFTTGLDAP-----------------RNLRR-------- 900

Qy        488 TPPAVKQTISNISGFNETCLRWRSIKTADMEEMYLFHIWGQRWYQKEFA-------QEMT 540
Db        901 ----VSQTDNSIT------LEWRNGKAAIDS-----------YRIKYAPISGGDHAEVD 938

Qy        541 FNISSSSRDPEVCLDLRPGTNYNVSLRALSSEL---PVVISLTTQITEP 586
Db        939 VPKSQQATTKTTLTGLRPGTEYGIGVSAVKEDKESNPATINAATELDTP 987


Search completed: February  6, 2004, 17:40:10
Job time : 24 secs
```